7

8250

THE

PHYSICAL SOCIETY

OF

LONDON.

ON THE

RELATIVITY THEORY OF GRAVITATION.

BY

A. S. EDDINGTON, M.A., M.Sc., F.R.S.
Plum[ian Pr]ofessor of Astronomy and Experimental Philosophy, Cambridge.

Price to Non-Fellows, **6s.** *net, post free* **6s. 3d.**
Bound in cloth, **8s. 6d.,** *post free* **8s. 9d.**

LONDON:
FLEETWAY PRESS, LTD.,
1, 2 AND 3, SALISBURY COURT, FLEET STREET.

1920.

CONTENTS.

CHAPTER I.

CHAPTER II.

CHAPTER III.

CHAPTER IV.

CHAPTER V.

CHAPTER VI.

CHAPTER VII.

CHAPTER VIII.

PREFACE TO FIRST EDITION.

THE relativity theory of gravitation in its complete form was published by Einstein in November 1915. Whether the theory ultimately proves to be correct or not, it claims attention as one of the most beautiful examples of the power of general mathematical reasoning. The nearest parallel to it is found in the applications of the second law of thermo-dynamics, in which remarkable conclusions are deduced from a single principle without any inquiry into the mechanism of the phenomena; similarly, if the principle of equivalence is accepted, it is possible to stride over the difficulties due to ignorance of the nature of gravitation and arrive directly at physical results. Einstein's theory has been successful in explaining the celebrated astronomical discordance of the motion of the perihelion of Mercury, without introducing any arbitrary constant; there is no trace of forced agreement about this prediction. It further leads to interesting conclusions with regard to the deflection of light by a gravitational field, and the displacement of spectral lines on the sun, which may be tested by experiment.

The arrangement of this Report is guided by the object of reaching the theory of these crucial phenomena as directly as possible. To make the treatment rather more elementary, use of the principle of least action and Hamiltonian methods has been avoided; and the brief account of these in Chapter VII. is merely added for completeness. Similarly, the equations of electro-dynamics are not used in the main part of the Report. Owing to the historical tradition, there is an undue tendency to connect the principle of relativity with the electrical theory of light and matter, and it seems well to emphasize its independence. The main difficulty of this subject is that it requires a special mathematical calculus, which, though not difficult to understand, needs time and practice to use with facility. In the older theory of relativity the somewhat forbidding vector products and vector operators

constantly appear. Happily this can now be avoided altogether; but in its place we use the absolute differential calculus of Ricci and Levi-Civita. This is developed *ab initio* so far as required in Chapter III. Attention must be called to the remark on notation in §19, which concerns almost all the subsequent formulæ.

Extensive use has been made of the following Papers, which in some places have been followed rather closely :—

A. EINSTEIN.—Die Grundlage der allgemeinen Relitivitätstheorie. "Annalen der Physik," XLIX., p. 769 (1916).

W. DE SITTER.—On Einstein's Theory of Gravitation and its Astronomical Consequences. "Monthly Notices of the Royal Astr. Soc.," LXXVI., p. 699 (1916); LXXVII., p. 155 (1916); LXXVIII., p. 3 (1917).

I am especially indebted to Prof. de Sitter, who has kindly read the proof-sheets of this Report.

The principal deviations in the present treatment of the subject will be found in Chapter VI. I have ventured to modify the enunciation of the principle of equivalence in §27 in order to give a precise criterion for the cases in which it is assumed to apply.

Other important Papers on the subject, most of which have been drawn on to some extent, are :—

H. HILBERT.—Die Grundlagen der Physik, "Göttingen Nachrichten," 1915, Nov. 20.

H. A. LORENTZ.—On Einstein's Theory of Gravitation, "Proc. Amsterdam Acad.," XIX., p. 1341 (1917).

J. DRÖSTE.—The Field of n moving centres on Einstein's Theory, "Proc. Amsterdam Acad.," XIX., p. 447 (1916).

A. EINSTEIN.—Kosmologische Betrachtungen zur allgemeinen Relitivitätstheorie, "Berlin Sitzungsber.," 1917, Feb. 8. Ueber Gravitationswellen, *ibid*, 1918, Feb. 14.

K. SCHWARZSCHILD.—Ueber das Gravitationsfeld eines Massenpunktes nach der Einstein'schen Theorie, "Berlin Sitzungsber.," 1916, Feb. 3.

T. LEVI-CIVITA.—Statica Einsteiniana, "Rendiconti dei Lincei," 1917, p. 458.

A. PALATINI.—Lo Spostamento del Perielio di Mercurio, "Nuovo Cimento," 1917, July.

The last two Papers avoid much of the heavy algebra, but claim a rather extensive knowledge of differential geometry.

The older theory of relativity, briefly surveyed in the first chapter, is fully treated in the well-known text-books of L.

Silberstein (Macmillan & Co.) and E. Cunningham (Camb. Univ. Press). A useful review of the mathematical theory of Chapter III., giving a fuller account from the standpoint of the pure mathematician, will be found in "Cambridge Mathematical Tracts," No. 9, by J. E. Wright. Finally, for those who wish to learn more of the outstanding discrepancies between astronomical observation and gravitational theory, the following references may be given :—

W. de Sitter.—The Secular Variations of the Elements of the Four Inner Planets, "Observatory," XXXVI., p. 296.

E. W. Brown.—The Problems of the Moon's Motion, "Observatory," XXXVII., p. 206.

H. Glauert.—The Rotation of the Earth, "Monthly Notices of the Royal Astr. Soc.," LXXV., p. 489.

PREFACE TO SECOND EDITION.

THE advances made in the eighteen months since this Report was written do not seem to call for any modification in the general treatment. Perhaps the most notable event is the verification of Einstein's prediction as to the deflection of a ray of light by the sun's gravitational field. This was tested at the total eclipse of May 29, 1919, at two stations independently, by expeditions sent out by the Royal and Royal Astronomical Societies jointly, under the superintendence of the Astronomer Royal. The deflection, reduced to the sun's limb should be $1''\cdot75$ on the relativity theory, and $0''\cdot87$ (or possibly zero) according to previous theories. At Principe, where the observations were very much interfered with by cloud, the value $1''\cdot61$ was obtained, with a probable error of $0''\cdot3$; the accuracy appears to be sufficient to indicate fairly decisively Einstein's value. At Sobral, where a clear sky prevailed, the observed value was $1''\cdot98$; the accordance of results derived from right ascensions and declinations, respectively, and the agreement of the displacements of individual stars with the theoretical law demonstrate in a particularly satisfactory manner the trustworthiness of the observations at this station. The full results will be published in a Paper by Sir F. W. Dyson, A. S. Eddington, and C. Davidson in the Philosophical Transactions of the Royal Society.

The test of the displacement of the Fraunhofer lines to the red stands where it did, and we still think that judgment must be reserved. In view of the possibility of a failure in this test, it is of interest to consider exactly what part of the theory can now be considered to rest on a definitely experimental basis. I think it may now be stated that Einstein's law of gravitation is definitely established by observation in the following form :—

Every particle and light-pulse moves so that the integral of ds between two points on its track is stationary, where (equation (28·8))

$$ds^2 = -(1-2m/r)^{-1}dr^2 - r^2d\theta^2 - r^2\sin^2\theta d\phi^2 + (1-2m/r)dt^2$$

in appropriate polar co-ordinates, the co-efficient of dr^2 being verified to the order m/r, and the co-efficient of dt^2 to the order m^2/r^2. This is checked for high speeds by the deflection of light, and for comparatively low speeds by the motion of perihelion of Mercury, so that unless the true law is of a kind much more complicated than we have allowed for, our expression cannot well be in error.

Accepting Einstein's law in this form, the properties of invariance for transformations of co-ordinates follow, and we reach the conclusion that the intermediary quantity ds (to which as yet we have assigned no physical interpretation) is an invariant, that is to say it has some absolute significance in external nature.

Einstein's theory (as distinct from his *law* of gravitation) gives a physical interpretation to ds, as a quantity that can be measured with material scales and clocks. It is this interpretation which the observation of the Fraunhofer lines should test. The quantity ds is an ideal measure of space and time; and it is possible that we have not yet reached finality as to the right way of realising the ideal practically. It is a fair prediction that an atomic vibration will register ds like an ideal clock; and it is difficult to see how this can be avoided unless the equations of vibration of an atom involve the Riemann-Christoffel tensor. But, if the test fails, the logical conclusion would seem to be that we know less about the conditions of atomic vibration than we thought we did.

A very notable extension of the theory to include electromagnetic forces and gravitational forces in one geometrical scheme has been given by Prof. H. Weyl in two Papers—

Berlin, Sitzungsberichte, 1918, May 30.
Annalen der Physik, Bd. 59, p. 101.

In Einstein's theory it is assumed that the interval ds has an absolute value, so that two intervals at different points of the world can be immediately compared. In practice the comparison must take place by steps along an intermediate path; for example, by moving a material measuring rod from one point to the other continuously along some path. It is possible that the result of the comparison may not be

independent of the path followed, and Weyl considers the electromagnetic field to be the manifestation of this inconsistency. This leads to a very beautiful generalized geometry of the world, in which the electromagnetic field appears as the sign of non-integrability of gauge, and the gravitational field as the sign of non-integrability of direction. The theory has important consequences though it has not suggested any experimental test. It may be added that it appears to favour Einstein's view of the curvature of space, which has been treated, perhaps too unsympathetically, in Chapter VIII.

The writer holds the view that the fundamental equations of gravitation (35·8), which on this theory are the sole basis of mechanics, should be regarded as a definition of matter rather than as a law of nature. We need not suppose that the gravitational field has *in vacuo* some innate tendency to arrange itself according to the law $G_{\mu\nu}=0$; we should rather say that in regions of the world where this state happens to exist we perceive emptiness ; and where the equations fail, the failure of the equations is itself the cause of our perception of matter. Matter does not cause the curvature (G) of space-time ; it *is* the curvature. Just as light does not cause electromagnetic oscillations ; it is the oscillations. This point of view is developed in a Paper which will appear shortly in "Mind."

Finally, a word may be added for those who find a difficulty in the combination of space and time into a static four-dimensional world, in which events do not "happen"—they are just there, and we come across them successively in our exploration. "Surely there is a difference between the irrevocable past and the open future, different in quality from the arbitrary distinction of right and left." We agree entirely ; but this difference, whatever it is, does not enter into the determinate equations of physics. For physics, the future is $+t$ and the past $-t$, just as right is $+x$ and left $-x$. If we change the place of one particle in our problem we alter the past as well as the future, in contrast to what appears to be the ordinary experience of life, that our interference will alter the future but not the past. The static four-dimensional representation may thus be not completely adequate, but it suffices for all that comes within the purview of physics.

December, 1919.

CHAPTER I.

THE RESTRICTED PRINCIPLE OF RELATIVITY.

1. In 1887 the famous Michelson-Morley experiment was performed with the object of detecting the earth's motion through the æther. The principle of the experiment may be illustrated by considering a swimmer in a river. It is easily realized that it takes longer to swim to a point 50 yards up-stream and back than to a point 50 yards across-stream and back. If the earth is moving through the æther there is a river of æther flowing through the laboratory, and a wave of light may be compared to a swimmer travelling with constant velocity relative to the current. If, then, we divide a beam of light into two parts, and send one half swimming up the stream for a certain distance and then (by a mirror) back to the starting point, and send the other half an equal distance across-stream and back, the across-stream beam should arrive back first.

Let the æther be flowing relative to the apparatus with velocity u in the direction Ox (Fig. 1); and let OA, OB be the two arms of the apparatus of equal length a, OA being placed up-stream. Let v be the velocity of light. The time for the double journey along OA and back is

$$t_1=\frac{a}{v-u}+\frac{a}{v+u}=\frac{2av}{v^2-u^2}=\frac{2a}{v}\beta^2 \quad . \quad . \quad . \qquad (1\cdot 1)$$

where $\beta=(1-u^2/v^2)^{-\frac{1}{2}}$, a factor greater than unity.

For the transverse journey the light must have a component velocity u up-stream (relative to the æther) in order to avoid being carried below OB; and, since its total velocity is v, its component across-stream must be $\sqrt{(v^2-u^2)}$. The time for the double journey OB is accordingly

$$t_2=\frac{2a}{\sqrt{(v^2-u^2)}}=\frac{2a}{v}\beta, \qquad (1\cdot 2)$$

so that $t_1>t_2$.

But when the experiment was tried, it was found *that both* parts of the beam took the same time, as tested by the interference bands produced. It would seem that OA and OB could not really have been of the same length; and if OB was of length a_1, OA must have been of length a_1/β. The apparatus was now rotated through 90°, so that OB became the up-stream arm. The time for the two journeys was again the same, so that OB must now be the shorter arm. The plain meaning of the experiment is that both arms have a length a_1 when placed along Oy, and automatically contract to a length a_1/β when placed along Ox. This explanation was first given by FitzGerald.

It is not known how much the earth's motion through the æther amounts to; but at some time during the year it must

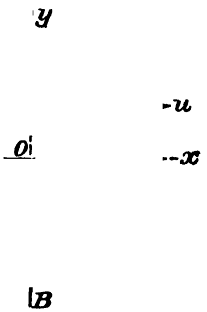

FIG. 1.

be at least 30 km. per sec., since the earth's velocity changes by 60 km. per sec. between opposite seasons. The experiment would have detected a velocity much smaller than this (about 3 km. per sec.), if it were not for the compensating contraction of the arms of the apparatus. By experimenting at different times of the year with different orientations the existence of the contraction has been fully demonstrated. It has been shown that it is independent of the material used for the arms, and the contraction is in all cases measured by the ratio $\beta=(1-u^2/v^2)^{-\frac{1}{2}}$.

It is now known that this contraction fits in well with the electrical theory of matter, and may be attributed to changes in the electromagnetic forces between the particles which determine the equilibrium form of a so-called rigid body. This universal property of matter is therefore not so mysterious

as it at first seemed; and we shall not here discuss the unsuccessful attempts at alternative explanations of the Michelson-Morley experiment, *e.g.*, by assuming a convection of the æther by the earth.

2. The Michelson-Morley experiment has thus unexpectedly failed to measure our motion through the æther, and many other ingenious experiments have failed in like manner. So far as we can test, the earth's motion makes absolutely no difference in the observed phenomena; and we shall not be led into any contradiction with observation if we assign to the earth any velocity through the æther that we please. It is interesting to trace in a general way how this can happen. Let us assign to the earth a velocity of 161,000 miles a second, say, in a vertical direction. With this speed $\beta=2$, and the contraction is one-half. A rod 6 feet long when horizontal contracts to 3 feet when placed vertically. Yet we never notice the change. If the standard yard-measure is brought to measure it, the rod will still be found to measure two yards; but then the yard-measure experiences the same contraction when placed alongside, and represents only half-a-yard in that position. It might be thought that we ought to see the change of length when the rod is rotated. But what we perceive is an image of the rod on the retina of the eye; we think that the image occupies the same space of retina in both positions; but our retina has contracted in the vertical direction without our knowing it, and our estimates of length in that direction are double what they should be. Similarly with other tests. We might introduce electrical and optical tests, in which the cause of the compensation is more difficult to trace; but, in fact, they all fail. The universal nature of the change makes it impossible to perceive any change at all.

3. This discussion leads us to consider more carefully what is meant by the *length* of an object, and the *space* which we consider it to occupy. To the physicist, space means simply a scaffolding of reference, in which the mind instinctively locates the phenomena of nature. Our present point of view assumes that there is a "real" or "absolute" scaffolding, in which a material body moving with the earth changes its length according as it is oriented in one direction or another. On the other hand, the human race (and its predecessors) have conceived and used a different scaffolding—the space of appearance—in which a material body moving with the earth does not change length as its orientation alters. It often

happens that a primitive conception is ambiguous, and has to be re-defined when adopted for scientific purposes; but there is little justification for doing this in the case of space. Firstly, the space of appearance is perfectly suitable for scientific purposes, since we have just seen that it is impossible to detect experimentally that it is not the absolute space Secondly, so long as we cannot detect our motion through the æther, we do not know how to convert our observations so as to express them in terms of absolute space. Thirdly, for all we know, our velocity through the æther may be so great that the absolute space and the space of appearance do not even approximately correspond; thus we might be revolutionising rather than re-defining the common conception of space.

It will therefore be considered legitimate to use the words "space" and "length" with their current significance. A rigid body on the earth is generally considered not to change length when its direction is altered, and by this property we block out a scaffolding of reference for our measures and locate objects in *our* space—the space of appearance. But we have learnt one important thing. Our space is not absolute; it is determined by our motion. If we transfer ourselves to the star Arcturus, which is moving relatively to us with a speed of more than 300 km. per sec., our space will not suit it, since it was designed to eliminate our own contraction effects. The contraction ratio β must be different for Arcturus; and the space surveyed with a material yard-measure carried on Arcturus will differ slightly from the space surveyed with the same yard-measure on the earth. It may also be noted that there is a slight difference in our own space in summer and winter (owing to the change of the earth's motion), and this may have to be taken into account in some applications.

Accordingly by "space" we shall mean the space of appearance for the observer considered. It becomes definite when we specify the motion of the observer. In particular, if the observer is at rest in the æther, the corresponding space is what we have hitherto called the "absolute space."

The possibility of different observers using different spaces may be illustrated by considering the question, What is a circle? Suppose a circle is drawn on paper in the usual way with a pair of compasses. An observer S, who believes the paper to be moving through the æther with a great velocity,

must, in accordance with the Michelson-Morley experiment, suppose that the distance between the points of the compasses changed as the curve was described; he will therefore deem the curve to be an ellipse. Another observer S', who believes the paper to be at rest in the æther, will deem it to be a circle. There is no experimental means of finding out which is right in his hypothesis. We have, therefore, to admit that the same curve may be arbitrarily regarded as an ellipse or as a circle. That illustrates our meaning when we say that S and S' use different spaces, the curve being an ellipse in one space and a circle in the other.

4. The failure of all experimental tests to decide whether the space of S or of S' is the more fundamental is summed up in the restricted Principle of Relativity. This asserts that *it is impossible by any conceivable experiment to detect uniform motion through the æther.* This generalisation is based on a great amount of experimental evidence, which is fully discussed in text-books on the older theory of relativity. Here it is perhaps sufficient to state that experimental confirmation appears to be sufficient,* except in regard to the question whether gravitation falls within the scope of the principle. We shall assume that the principle is true universally.

Let x', y', z', be the co-ordinates of a point in the space of an observer S'; and let x, y, z be the co-ordinates of the same point in the space of an observer S at rest in the æther. Let S' move relatively to S with velocity u in the direction Ox. S', using his own space, has no knowledge of his motion through the æther, and he makes all his theoretical calculations as though he were at rest; from what has been already said, he will not discover any contradiction with observation.

According to ordinary kinematics the relation between the co-ordinates and the times (t', t) in the two systems would be

$$x'=x-ut, \quad y'=y, \quad z'=z, \quad t'=t \quad . \quad . \quad . \quad (4\cdot 1)$$

But the first of these must be modified, because in the x-direction S''s standard of length is contracted in the ratio $1/\beta$. The equation becomes

$$x'=\beta(x-ut). \quad . \quad . \quad . \quad . \quad . \quad . \quad (4\cdot 15)$$

In order to satisfy the principle of relativity, it appears that the time t' used by S' must differ from the time t used by S.

* *i.e.*, sufficient to assert the *universality*, not necessarily the perfect *accuracy*, of the principle.

We shall suppose that both observers use the same value for the velocity of light; this is merely a matter of co-ordinating their units, the significance of which will be considered in the next paragraph. Let S' observe the time t' taken for the double journey $OB=2a_1$, in Fig. 1. It must agree with his calculated time, which is, of course, $2a_1/v$. Thus

$$t'=2a_1/v.$$

But in (1·2), when we were using S's co-ordinates, we found the time to be

$$t=2a_1\beta/v.$$

Hence

$$t=\beta t'.$$

This also fits the double journey OA. S', unaware of his motion, does not allow for any contraction, and calculates the time for the double journey as

$$t'=2a_1/v.$$

But S recognises the contraction, and considers the distance travelled to be $2a_1/\beta$. Hence calculating as in (1·1), he makes the time to be

$$t=\frac{2a_1}{\beta v}\beta^2,$$

so that again $$t=\beta t'.$$

Accordingly S' must use a unit of time longer than that of S in the ratio β; otherwise he would find a discrepancy between observation and calculation.

There is another difference in time-measurement involved. According to S, the light completes the half-journey OA in a time $\frac{a_1/\beta}{v-u}$ in S's units, or $\frac{a_1/\beta^2}{v-u}$ in S''s units of time. But

$$\frac{a_1}{\beta^2(v-u)}=\frac{a_1(v+u)}{v^2}=\frac{a_1}{v}+\frac{a_1u}{v^2}.$$

But the difference in the time of leaving O and reaching A must be deemed by S' to be a_1/v; he must therefore set his clock at A a_1u/v^2 slow compared with the clock at O. He has no idea that it is slow; he has attempted to adjust the two clocks together. But his determination of simultaneity of events at O and A differs from that of S, because he allows a different correction for the time of transit of the light.

Including both these differences, we see that the relation between the times adopted by S and S' is

$$t=\beta\left(t'+\frac{ux'}{v^2}\right).$$

Substituting this value of t in (4·15) we obtain after an easy reduction

$$x=\beta(x'+ut').$$

Collecting together our results, we have the formulæ of transformation

$$x=\beta(x'+ut'),\quad y=y',\quad z=z',\quad t=\beta\left(t'+\frac{ux'}{v^2}\right). \quad (4·2)$$

By the principle of relativity nothing is altered if S is in motion relative to the æther; so the relations (4·2) must hold between the spaces and times of *any* two observers having relative velocity u.

By solving (4·2) for x', y', z', t', we obtain the reciprocal relations

$$x'=\beta(x-ut),\quad y'=y,\quad z'=z,\quad t'=\beta\left(t-\frac{ux}{v^2}\right) \quad . \quad . \quad (4·3)$$

These might have been written down immediately, because interchanging S and S' is equivalent to reversing the sign of u; but it will be seen later that the verification by direct solution of (4·2) is important.

5. We have supposed that S and S' adopt the same measure for the velocity of light; this was in order to secure that the units of velocity used by S and S' correspond. It is no use for S to describe his experiences to S' in terms of units which are outside the knowledge of the latter; but if S states that a velocity occurring in his experiment is a certain fraction of the velocity of light, S' will be able to compare that with his own experimental results. By the principle of relativity any other velocities occurring in their experiments under similar conditions will correspond; and, for example, we see from (4·2) and (4·3) that they will agree in calling their relative velocity $+u$ and $-u$ respectively.

Whilst this settles the consistency of the units of velocity used in (4·2) we have not yet secured that the units of length correspond. A description of Brobdingnag by a Brobdingnagian would not have mentioned the most striking feature of that country; it needed an intruding Gulliver to detect

the enormous scale of everything contained. And so we may ask whether a natural standard of length, say a hydrogen atom, at rest in S's system will be of the same size in terms of x, y, z, as a hydrogen atom at rest in S''s system in terms of x', y', z'. Clearly it will be misleading if we do not correlate the co-ordinates so as to satisfy this.

To allow for a possible non-correspondence of the units of length in (4·2) we can write the transformation more generally

$$kx=\beta(x'+ut'), \quad ky=y', \quad kz=z', \quad kt=\beta(t'+ux'/v^2) \quad . \quad (5{\cdot}2)$$

where k depends on the magnitude, but clearly not on the direction, of u.

But now applying (5·2) the reverse way, *i.e.*, regarding x, y, z, t as a system moving with velocity $-u$ relative to x', y', z', t', we shall have

$$kx'=\beta(x-ut), \quad ky'=y, \quad kz'=z, \quad kt'=\beta(t-ux/v^2) \quad . \quad (5{\cdot}3)$$

which is clearly inconsistent with (5·2) unless $k=1$. Hence (4·2) gives the only possible correspondence of the units of length.

We thus use the remarkable property of reciprocity possessed by (4·2) and (4·3), but not by (5·2) and (5·3), to fix the necessary correspondence of the units. The dimensions of a motionless hydrogen atom will now be the same in both systems; for, if not, we could find a system in which the dimensions were either a maximum or a minimum; and that system would give us an absolute standard from which we could measure absolute motion.

It is thus clear that S' will actually measure his space and time by the variables x', y', z', t' given by (4·3), if he sets about choosing his units in the same way that S did.

6. We have established the connection between the co-ordinates used by S and S' by reference to simple criteria. It is interesting to work out in detail the correspondence of the two systems for other and more complex phenomena, showing that the transformation always works consistently. But the standpoint of the principle of relativity rather discourages this procedure. Its view is that the indifference of all natural phenomena to an absolute translation is something immediately understandable, whilst the contractions and other complications entering into our description arise from our own perversity in not looking at Nature in a broad enough way. When a rod is started from rest into uniform motion, nothing whatever happens to the rod. We say that it contracts; but

length is not a property of the rod; it is a *relation* between the rod and the observer. Until the observer is specified the length of the rod is quite indeterminate. We ought always to remember that our experiments reveal only relations, and not properties inherent in individual objects; and then the correspondence of two systems, differing only in uniform motion, becomes axiomatic, so that laborious mathematical verifications are redundant. Human minds being what they are, that is a counsel of perfection, and we shall not follow it too strictly.

The only verification that is needed is to show that our fundamental laws of mechanics and electrodynamics are consistent with the principle of relativity. This will be done in connection with a much more general principle of relativity for mechanics in § 37, and for electrodynamics in § 45.

7. (*a*) As an illustration of the modification of ordinary views required by this theory, we may notice the law of composition of velocities. Consider a particle moving relative to S with velocity w along Ox, so that

$$\frac{dx}{dt}=w \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7{\cdot}1)$$

The velocity relative to S' will be

$$w'=\frac{dx'}{dt'}=\frac{\beta(dx-udt)}{\beta(dt-udx/v^2)} \quad \text{by } (4{\cdot}3),$$

$$=\frac{w-u}{1-uw/v^2} \quad \text{by } (7{\cdot}1) \quad . \quad . \quad . \quad . \quad (7{\cdot}2)$$

The velocity relative to S' is thus not $w-u$, as we should have assumed in ordinary mechanics.

It has been pointed out by Robb that the addition-law for motion in one dimension can be restored if we measure motion by the *rapidity*, $\tanh^{-1}(w/v)$, instead of by the velocity w. Equation (7·2) gives

$$\tanh^{-1}(w'/v)=\tanh^{-1}(w/v)-\tanh^{-1}(u/v) \quad . \quad . \quad (7{\cdot}3)$$

Since $\tanh^{-1}1=\infty$, the velocity of light corresponds to infinite rapidity, and we may compound any number of relative velocities less than that of light without obtaining a resultant greater than the velocity of light.

(*b*) To find the relation of the densities (σ=number of particles per unit volume) in the two systems, we can easily verify that the Jacobian $\partial(x', y', z', t')/\partial(x, y, z, t)=1$, so that

$$dx'dy'dz'dt'=dx\,dy\,dz\,dt \quad . \quad . \quad . \quad . \quad . \quad (7{\cdot}4)$$

But the number of particles in a particular element of volume cannot depend on the co-ordinates used to describe the element, hence

$$\sigma' dx' dy' dz' = \sigma dx\, dy\, dz. \quad . \quad . \quad . \quad . \quad (7\cdot5)$$

Hence
$$\frac{\sigma'}{\sigma} = \frac{dt'}{dt} = \beta\left(1 - \frac{uw}{v^2}\right) \quad . \quad . \quad . \quad . \quad . \quad (7\cdot6)$$

since $dx/dt = w$.

In particular, if $w=0$, so that σ is the density referred to axes moving with the matter,

$$\sigma' = \beta\sigma. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7\cdot65)$$

Since the mass of a particle may depend on its motion, we cannot assume that the ratio ρ'/ρ of the mass-density is the same as that of the distribution-density σ'/σ.

When the transformation (4·2) was first introduced in electro-dynamics by Larmor and Lorentz, t' was regarded as a fictitious time introduced for mathematical purposes, and it was scarcely realised that it was the actual measured time of the moving observer. Einstein in 1905 first showed that velocity and density would be estimated by the moving observer in the way given above, and thus removed the last discrepancy between the electrodynamical equations for the two systems.

(*c*) In order to find the change (if any) of mass with velocity, consider a body of mass m_1, m_1' (in the two systems of reference) moving with velocity w_1, w_1'. Let

$$\beta_1 = (1 - w_1^2/v^2)^{-\frac{1}{2}}, \quad \beta_1' = (1 - w_1'^2/v^2)^{-\frac{1}{2}}.$$

Working out β_1' by using (7·2), we easily find

$$\beta_1' w_1' = \beta\beta_1(w_1 - u) \quad . \quad . \quad . \quad . \quad . \quad (7\cdot71)$$

Let a number of bodies be moving in a straight line subject to the conservation of mass and momentum, *i.e.*,

$$\Sigma m_1 \quad \text{and} \quad \Sigma m_1 w_1 \quad \text{are conserved.}$$

Then, since u and β are constants,

$$\beta\Sigma m_1(w_1 - u) \text{ will be conserved.}$$

Therefore by (7·71)

$$\Sigma\frac{m_1\beta_1'}{\beta_1} w_1' \text{ is conserved.} \quad . \quad . \quad . \quad (7\cdot72)$$

But since momentum must be conserved for the observer S'

$$\Sigma m_1' w_1' \text{ is conserved} \quad . \quad . \quad . \quad . \quad (7\cdot73)$$

The results (7·72) and (7·73) will agree if

$$\frac{m_1}{\beta_1}=\frac{m_1'}{\beta_1'}=m_0, \text{ say,}$$

and it is easy to show that there is no other solution. Hence

$$m_1=m_0\beta_1=m_0(1-w_1^2/v^2)^{-\frac{1}{2}} \quad . \quad . \quad . \quad . \quad (7\cdot8)$$

where m_0 is constant and equal to the mass at rest. This is the law of dependence of mass on velocity.

Neglecting w_1^4/v^4, we have

$$m_1=m_0+(\tfrac{1}{2}m_0w_1^2)/v^2 \quad . \quad . \quad . \quad . \quad (7\cdot85)$$

so that we may regard the mass as made up of a constant mass m_0 belonging to the particle, together with a mass proportional to, and presumably belonging to, the kinetic energy. If we choose units so that the velocity of light is unity, the mass of the energy is the same as the energy, and the distinction between energy and mass is obliterated. Accordingly m_0 is also regarded as a form of energy. (It is usually identified mainly with the electrostatic energy of the electrons forming the body.)

Since the conservation of mass now implies the conservation of energy we have to restrict the reactions between the bodies in the foregoing discussion to perfectly elastic impacts. Other interactions would require a more general treatment ; in fact, if the energy is not conserved, the momentum is not perfectly conserved, because the disappearing energy has mass and therefore carries off momentum.

In this discussion we are justified in pressing the laws of conservation of mass and momentum to the utmost limit as holding with absolute accuracy, since the definition and measurement of mass (inertia) rests on these laws,* and unless we have an accurate definition it is meaningless to investigate change of mass. In astronomy, however, the masses of heavenly bodies are measured by their gravitational effects ; naturally we cannot legitimately apply (7·8) to *gravitational mass* without a full discussion of the law of gravitation.

It should be noticed that this change of mass with velocity is in no way dependent on the electrical theory of matter.

* The mass here discussed is sometimes called the "transverse mass." The so-called longitudinal mass is of no theoretical importance ; it is not conserved, it does not enter into the expression for the momentum or energy, and it has no connection with gravitation.

(*d*) To find the transformation of mass-density, ρ, we have

$$\frac{\rho'}{\rho}=\frac{\sigma' m'}{\sigma m}=\frac{\sigma'\beta_1'}{\sigma\,\beta_1}=\frac{\beta_1'\beta}{\beta_1}\left(1-\frac{uw_1}{v^2}\right) \text{ by (7·6),}$$

which becomes by (7·71)

$$\frac{\rho'}{\rho}=\left(\frac{\beta_1'}{\beta_1}\right)^2 \qquad (7\cdot91)$$

In particular, if ρ_0 is the density in natural measure, *i.e.*, referred to axes moving with the matter, ρ the density referred to axes with respect to which the matter has a velocity u,

$$\rho=\beta^2\rho_0 \; . \; . \; . \; . \; . \; . \; . \; . \qquad (7\cdot92)$$

8. Of late years the domain of the electromagnetic theory has been extended, so that most natural phenomena are now attributed to electrical actions. The relativity theory does not presuppose an electromagnetic theory either of matter or of light; but, if we accept the latter theories, it becomes possible to state exactly the points on which experimental evidence is required in order to establish our hypothesis. The experimental laws of electromagnetism are summed up in Maxwell's equations; and in so far as these cover the phenomena, the complete equivalence of the sequence of events in a fixed system described in terms of x, y, z, t, and a moving system described in terms of x', y', z', t', has been established analytically. So far as is known, only three kinds of force are outside the scope of Maxwell's equations.

(1) The forces which constrain the size and shape of an electron are not recognised electromagnetic forces. Fortunately the properties of an electron at rest and in extremely rapid motion can be studied experimentally, and it is believed that they change in the way required by relativity.

(2) The phenomena of Quanta appear to obey laws outside the scope of Maxwell's equations. Theoretically these laws fit in admirably with relativity, since Planck's fundamental unit of action is found to be unaltered by the choice of axes. But on the experimental side, evidence of the relativity of phenomena involving quantum relations has not yet been produced. This is particularly unfortunate, because the vibration of an atom depends on quantum relations; and it is practically essential to the relativity theory that an atom (acting as a natural clock) should keep the time appropriate to the axes chosen.

(3) Gravitation is outside the electromagnetic scheme. The Michelson-Morley experiment is necessarily confined to solids of laboratory dimensions, in which internal gravitation has no appreciable influence. There is, therefore, no experimental proof that a body such as the earth, whose figure is determined mainly by gravitation, will undergo the theoretical contraction owing to motion. The most direct evidence that gravitation conforms to relativity comes from a discussion by Lodge* of the effect of the sun's motion through the æther on the perihelia and eccentricities of the inner planets. If gravitation is outside the relativity theory (the Newtonian law holding unmodified) a solar motion of 10 km. per sec. would produce perturbations in the eccentricities and perihelia of the earth and Venus, which could probably be detected by observation. The absence of these perturbations seems to show that gravitation must conform to relativity, unless, indeed, the sun happens to be nearly at rest in the æther. If we confine attention to our local stellar system the average stellar velocities are not so much greater than 10 km. per sec. as to render the latter alternative too improbable ; but the very high velocities found for the spiral nebulæ (which are thought to be distant stellar systems) makes it improbable that our local system should be so nearly at rest in the æther.

* "Phil. Mag.," February, 1918.

CHAPTER II

THE RELATIONS OF SPACE, TIME AND FORCE.

9. An interesting aspect of the transformation of the variables x, y, z, t to x', y', z', t' has been brought out by Minkowski. We consider them as co-ordinates in a four-dimensional continuum of space and time. Choose the units of space and time so that the velocity of light is unity, and set

$$t=i\tau, \quad \text{where } i=\sqrt{-1}.$$

The equations of transformation (4·2) become

$$x=\beta(x'+iu\tau'), \quad y=y', \quad z=z', \quad \tau=\beta(\tau'-iux') \quad . \quad (9\cdot1)$$
$$\beta=(1-u^2)^{-\frac{1}{2}}.$$

Let $u=i\tan\theta$, so that θ is an imaginary angle. Then $\beta=\cos\theta$, and (9·1) becomes

$$x=x'\cos\theta-\tau'\sin\theta, \quad y=y', \quad z=z', \quad \tau=\tau'\cos\theta+x'\sin\theta \quad . \quad (9\cdot2)$$

Thus the transformation is simply a rotation of the axes of co-ordinates through an imaginary angle θ in the plane of $x\tau$.

We know that the orientation chosen for the space-axes, x, y, z, makes no difference in Newtonian mechanics. The principle of relativity extends this so as to include the axis τ. The continuum ormed o' space and imaginary time is perfectly isotropic ; the resolution into space and time separately, which depends on the motion of the observer, corresponds to the arbitrary orientation in it of a set of rectangular axes.

10. From this point of view the strange conspiracy of the forces of Nature to prevent the detection of our absolute motion disappears There is no conspiracy of concealment, because there is nothing to conceal. The continuum being isotropic, there is no orientation more fundamental than any other ; we cannot pick out any direction as the absolute time any more than we can pick out a direction in space as the absolute vertical. Up-and-down, right-and-left, backwards-

and-forwards, sooner-and-later,* equally express relations to some particular observer, and have no absolute significance. In Minkowski's famous words, "Henceforth Space and Time in themselves vanish to shadows, and only a kind of union of the two preserves an independent existence."

The scientific basis of the idea that some fundamental division into space and time exists was the conception of the æther as a material fluid, filling uniformly and isotropically a particular space. It now seems clear that the æther cannot have those material properties which would enable it to serve as a frame of reference. Its functions seem to be limited to those summed up in the old description "the nominative of the verb 'to undulate.'"

Unfortunately the simplicity of this conception of the four-dimensional continuum is only formal; and natural phenomena make a discrimination between τ and the other variables by relating themselves to an imaginary τ, which we call the time. In natural variables, $x, y, z, t,$ this view of the transformation as a rotation of axes becomes concealed.†

11. In the four-dimensional continuum the interval δs between two point-events is given by

$$-\delta s^2=\delta x^2+\delta y^2+\delta z^2+\delta\tau^2 \quad . \quad . \quad . \quad (11{\cdot}1)$$

which is unaffected by any rotation of the axes, and is therefore invariant for all observers. The minus sign given to δs^2 is an arbitrary convention, and the formula is simply the generalisation of the ordinary equation

$$\delta s^2=\delta x^2+\delta y^2+\delta z^2.$$

The fact that δs is measured consistently by all observers who would obtain discordant results for δx, δy, δz, $\delta\tau$ separately, is so important in our subsequent work that we shall consider the nature of the clock-scale needed for its measurement.

We have a scale AB divided into kilometres, say, and at each division is placed a clock also registering kilometres.

* This applies to imaginary time. With real time, events which (as usually happens) are separated by a greater interval in time than in space preserve the same order for all observers. But an event on the sun which we should describe as occurring 2 minutes later than an event on the earth might be described by another observer as 2 minutes earlier. (Both observers have corrected their observations for the light-time.)

† For a logical study of the properties of the continuum of space and real time reference may be made to A. A. Robb, "A Theory of Time and Space" (Camb. Univ. Press).

(The velocity of light being unity, a kilometre is also a unit of time $= \frac{1}{300000}$ sec.) When the clocks are correctly set and viewed from A, the sum of the readings of any clock and the division beside it is the same for all, since the scale-reading gives the correction for the time taken by light in travelling to A. This is shown in Fig. 2, where the clock-readings are given as though they were being viewed from A.

Now lay the scale in line with the two events; note the clock and scale-reading, t_1, σ_1, of the first event, and the corresponding readings t_2, σ_2, of the second event; then from (11·1)

$$\delta s^2=(t_2-t_1)^2-(\sigma_2-\sigma_1)^2 \quad . \quad . \quad . \quad . \quad (11{\cdot}2)$$

If the scale had been set in motion in the direction AB, $\sigma_2-\sigma_1$ would have been diminished, owing to the divisions having advanced to meet the second event. But the clocks would have been adjusted differently, because A is now

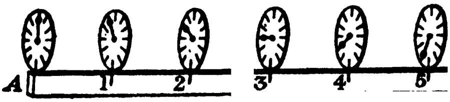

FIG. 2.

advancing to meet the light coming from any clock, and the clock would appear too fast (by the above rule) if it were not set back. There are other second-order corrections arising from the contraction of the scale and change of rate of the clocks owing to motion; but the net result is a perfect compensation, and δs^2 determined from (11·2) must be invariant, as already proved.

It is clear that the whole (restricted) principle of relativity is summed up in this invariance of δs, and it is possible to deduce the equation of transformation (4·2) and our other previous results by taking this as postulate.

When δs refers to the interval between two events in the history of a particular particle it has a special interpretation which deserves notice. If we choose axes moving with the particle, δx, δy, $\delta z=0$, so that $\delta s=\delta t$. Accordingly the variable s is called the "proper-time," *i.e.*, the time measured by a clock attached to the particle.

12. Up to the present we have discussed a particular type transformation of co-ordinates, viz., that corresponding to

a uniform motion of translation. We now enter on the theory of more general changes of co-ordinates.

The co-ordinates x, y, z, t of a particle trace a curve in four dimensions which is called the *world-line* of the particle. If we draw the world-lines of all the particles, light-waves and other entities, we obtain a complete history of the configurations of the Universe for all time. But such a history contains a great deal that is necessarily outside experience. All exact observations are records of coincidences of two entities in space and time, that is to say, records of intersections of world-lines.

It is easy to see that this is the case in laboratory experiments or astronomical observations. Electrical measurements, determinations of temperature, weight, pressure, &c., rest finally on the coincidence of some indicator with a division on a scale. Many of our rough observations depend on coincidences of light waves with elements of the retina, or the simultaneous impact of sound-waves on the ear. It is true that some of our external knowledge is not obviously of this character. We estimate the weight of a letter, balancing it in the hand; this is based on a muscular sensation having no immediate relation to time and space, but we fit this crude knowledge into the exact scheme of physics by comparing it with more accurate measures based on coincidences.

The observation that the world-lines of two particles intersect is a genuine addition to knowledge, since in general lines in space of three or four dimensions miss one another. We have to build up our conception of the location of objects in space and time from a large number of records of coincidences. It is clear that we have a great deal of liberty in drawing the world-lines, whilst satisfying all the intersections. Let us draw the world-lines in some admissible way, and imagine them embedded in a jelly. If the jelly is distorted in any way, the world-lines in their new courses will still agree with observation, because no intersection is created or destroyed.

Mathematically this can be expressed by saying that we may make any mathematical transformation of the co-ordinates. If we choose new co-ordinates x', y', z', t', which are any four independent functions of x, y, z, t, a coincidence in x, y, z, t will also be a coincidence in x', y', z', t', and *vice versa*. By locating objects in the space-time given by x', y', z', t', we do not alter the course of events. The events themselves do not presuppose any particular system of co-ordinates, and the

space-time scaffolding is something introduced arbitrarily by ourselves.

It is almost a truism to say that we may adopt any system of co-ordinates we please. We are accustomed to introduce curvilinear co-ordinates or moving axes without apology, whenever they simplify the problem. But there is one point not so generally recognised. Ordinarily when we use curvilinear co-ordinates we never allow ourselves to forget that they *are* curvilinear; it is a mathematical device, not a new space, that we adopt. Perhaps the only case in which we really take the new co-ordinates seriously is in the transformation to rotating axes; we then take account of the rotation by adding a fictitious centrifugal force to the equations, and thenceforth the rotation is quite put out of mind. From the standpoint of relativity, when we adopt new co-ordinates x', y', z', t', we shall adopt a corresponding new space, and think no more of the old space. For instance, a " straight line " in the new space will be given by a linear relation between x', y', z', t'.

The behaviour of natural objects will no doubt appear very odd when referred to a space other than that customarily used. So-called rigid bodies will change dimensions as they move; but we are prepared for that by our study of the Michelson-Morley contraction. Paths of moving particles will for no apparent reason deviate from the " straight line," but, accepting the definition of a force as that which changes a body's state of rest or motion, this must be attributed to a field of force inherent in the new space (cf. the centrifugal force). Light-rays will also be deflected, so that the field of force acts on light as well as on material particles; this is not altogether a novel idea, because a little reflection shows that the centrifugal force deflects light as well as matter—although optical problems are not usually treated in that way.

13. The laws of mechanics and electrodynamics are usually enunciated with respect to " unaccelerated rectangular axes," or, as they are often called, " Galilean axes." We cannot regard such axes as recognisable intuitively, and the only definition of them that can be given is that they are the axes with respect to which that particular form of the laws holds. It is part of the method of the present theory to restate the laws of Nature in a form not confined to Galilean co-ordinates, so that all systems of co-ordinates are regarded as on the same footing.

In unaccelerated rectangular co-ordinates the path of a particle is a straight line (apart from the influence of other matter, or the electromagnetic field). When we transform to other co-ordinates the path is no longer straight, *i.e.*, it is no longer given by a linear relation between the co-ordinates; and the bending of the path is attributable to a field of force which comes into existence in the new space. This field of force has the property that the deflection produced is independent of the nature of the body acted on, being a purely geometrical deformation. Now the same property is shared by the force of gravitation, the acceleration produced by a given gravitational field is independent of the nature or mass of the body acted on. This has led to the hypothesis that gravitation may be of essentially the same nature as the geometrical forces introduced by the choice of co-ordinates.

This hypothesis, which was put forward by Einstein, is called the Principle of Equivalence. It asserts that *a gravitational field of force is exactly equivalent to a field of force introduced by a transformation of the co-ordinates of reference, so hat by no possible experiment can we distinguish between them.*

In Jules Verne's story, "Round the Moon," three men are shot up in a projectile into space. The author describes their strange experiences when gravity vanishes at the neutral point between the earth and moon. Pedantic criticism of so delightful a book is detestable; yet perhaps we may point out that, for the inhabitants of the projectile, weight would vanish the moment they left the cannon's mouth. They and their projectile are falling freely all the time at the same rate, and they can feel no sensation of weight. They automatically adopt a new space, referred to the walls and fixtures of their projectile instead of to the earth. Their axes of reference are accelerated—falling towards the earth; and this transformation of axes introduces a field of force which just neutralises the gravitational field. But, whilst they could detect no gravitational field by ordinary tests, it is not obviously impossible for them to detect some effect by optical or electrical experiments. According to the principle of equivalence, however, no effect of any kind could be detected inside the projectile; the gravitational field cannot be differentiated from a transformation of co-ordinates, and therefore the same transformation which neutralises mechanical effects neutralises all other effects.

It will be seen that this principle of equivalence is a natural generalisation of the principle of relativity. An occupant of the projectile can only observe the *relations* of the bodies inside to himself and to each other. The supposed absolute acceleration of the projectile is just as irrelevant to the phenomena as a uniform translation is. The mathematical space-scaffolding of Galilean axes, from which we measure it, has no real significance. If the projectile were not allowed to fall, gravity would be detected—or rather the force of constraint which prevents the fall would be detected. I think it is literally true to say that we never feel the force of the earth's attraction on our bodies; what we do feel is the earth shoving against our feet.

14. A limitation of the Principle of Equivalence must be noticed. It is clear that we cannot transform away a natural gravitational field altogether. If we could, we should unconsciously make the transformation and adopt the new coordinates just as the inhabitants of the projectile did. They were concerned with a practically infinitesimal region, and for an infinitesimal region the gravitational force and the force due to a transformation correspond; but we cannot find any transformation which will remove the gravitational field throughout a finite region. It is like trying to paste a flat sheet of paper on a sphere, the paper can be applied at any point, but as you go away from the point you soon come to a misfit. For this reason it will be desirable to define the exact scope of the principle of equivalence. Up to what point are the properties of a gravitational field and a transformation field identical? And what properties does a gravitational field possess which cannot be imitated by a transformation? The impossibility of transforming away a gravitational field is, of course, an experimental property; so that, in spite of the principle of equivalence, there is at least one means of making an experimental distinction.

Space-time in which there is no gravitational field which cannot be transformed away is called *homaloidal*. In homaloidal space-time then, we can choose axes so that there is no field of force anywhere. Remembering that we have no means of defining axes except from the form of the laws of Nature referred to them, we should naturally take these axes as fundamental and name them "rectangular and unaccelerated." The dynamics of homaloidal space would not recognise the existence of gravitation. Our space is not like that, though

we believe that at great distances from all gravitating matter it tends towards this condition as a limit. The necessary limitation of the principle of equivalence turns on the number of consecutive points for which gravitational space-time agrees with homaloidal space-time; in other words, the equivalence will hold only up to a certain order of differential coefficients. Properties involving differential coefficients up to this order will be the same in the gravitational field as in a homaloidal field; whilst properties of the transformed field involving differential coefficients of higher order will not necessarily hold in the gravitational field

The determination of the order of the differential coefficients for which agreement is possible must be deferred to § 27. Meanwhile it may be noted that we can always choose axes for which the field at a given *point* vanishes—viz., take rectangular axes moving with the acceleration at that point. In that case we are said to use "natural measure."

15. At a point of space where there is no field of force the observer's clock-scale, if unconstrained, will be either at rest or in uniform motion. We have seen that the measured interval, δs, between two events is independent of uniform motion, and hence a unique value of δs is determined by the measures.

Using rectangular co-ordinates, the relation between an infinitesimal *measured* interval ds and the *inferred* co-ordinates of the event is (11·1).

$$ds^2 = -dx^2 - dy^2 - dz^2 + dt^2 \quad . \quad . \quad . \quad . \quad (15\cdot1)$$

Introduce new co-ordinates x_1, x_2, x_3, x_4, which are any functions of x, y, z, t given by

$$x = f_1(x_1, x_2, x_3, x_4), \quad y = f_2(x_1, x_2, x_3, x_4), \ \&c.$$

Then $$dx = \frac{\partial f_1}{\partial x_1}dx_1 + \frac{\partial f_1}{\partial x_2}dx_2 + \frac{\partial f_1}{\partial x_3}dx_3 + \frac{\partial f_1}{\partial x_4}dx_4, \ \&c. \quad . \quad (15\cdot2)$$

Substituting (15·2) on the right-hand side of (15·1), we obtain a general quadratic function of the infinitesimals, which may be written,

$$ds^2 = g_{11}dx_1^2 + g_{22}dx_2^2 + g_{33}dx_3^2 + g_{44}dx_4^2 + 2g_{12}dx_1dx_2 + 2g_{13}dx_1dx_3 + 2g_{14}dx_1dx_4 + 2g_{23}dx_2dx_3 + 2g_{24}dx_2dx_4 + 2g_{34}dx_3dx_4 \quad . \quad (15\cdot3)$$

where the g's are functions of the co-ordinates, depending on the transformation.

As an illustration we may take the transformation to rotating axes

$$\left.\begin{aligned} x&=x_1 \cos \omega x_4-x_2 \sin \omega x_4 \\ y&=x_1 \sin \omega x_4+x_2 \cos \omega x_4 \\ z&=x_3 \\ t&=x_4 \end{aligned}\right\} \quad . \quad . \quad . \quad (15\cdot4)$$

Whence

$$dx= \cos \omega x_4.dx_1- \sin \omega x_4.dx_2-\omega(x_1 \sin \omega x_4+x_2 \cos \omega x_4)dx_4,$$
$$dy= \sin \omega x_4.dx_1+ \cos \omega x_4.dx_2+\omega(x_1 \cos \omega x_4-x_2 \sin \omega x_4)dx_4,$$
$$dz=dx_3, \; dt=dx_4.$$

Substituting in (15·1)

$$ds^2= -dx_1^2 -dx_2^2 -dx_3^2+(1-\omega^2(x_1^2+x_2^2))dx_4^2+2\omega x_2 dx_1 dx_4 -2\omega x_1 dx_2 dx_4 \quad . \quad . \quad . \quad . \quad . \quad . \quad (15\cdot5)$$

By comparing this with (15·3) we obtain the values of the g's for this system of co-ordinates.

16. These values of the g's express the metrical properties of the space that is being used. But the observer has no immediate perception of them as properties of space. He does not realise that there is anything geometrically unnatural about axes rotating with the earth, but he perceives a field of centrifugal force. Experiments, such as Foucault's pendulum and the gyro-compass, designed to exhibit the absolute rotation of the earth, are more naturally interpreted as detecting this field of force.

Thus the coefficients g_{11}, &c., can be taken as specifying a field of force. That they are *sufficient* to define it completely may be seen from the following consideration. The world-line of a particle under no forces is a straight line in the system x, y, z, t, and its equation may be written in the form

$$\int ds \text{ is stationary;} \quad . \quad . \quad . \quad . \quad . \quad (16\cdot1)$$

but in this form the equation is independent of the choice of co-ordinates, and applies to all systems. If we choose new co-ordinates, the world-line given by (16·1) becomes curved and the curvature is attributed to the field of force introduced; but clearly the curvature of the path can only depend on the expression for ds in the new co-ordinates, *i.e.*, on the g's. Thus the force is completely defined by the g's.

It will be noticed that in (15·5)

$$g_{44}=1-2\Omega, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16\cdot2)$$

where $\Omega=\frac{1}{2}\omega^2(x_1^2+x_2^2)$=the potential of the centrifugal force.

Thus g_{44} can be regarded as a potential; and by analogy all the coefficients are regarded as components of a generalised potential of the field of force.

According to the principle of equivalence it must also be possible to specify a gravitational field by a set of values of the g's. It will be our object to find the differential equations satisfied by the g's representing a gravitational field. These differential equations for the generalised potential will express the law of gravitation, just as the Newtonian law is expressed by $\nabla^2\varphi=0$.

The double aspect of these coefficients, g_{11}, &c., should be noted. (1) They express the metrical properties of the co-ordinates. This is the official standpoint of the principle of relativity, which scarcely recognises the term "force." (2) They express the potentials of a field of force. This is the unofficial interpretation which we use when we want to translate our results in terms of more familiar conceptions.

Although we deny absolute space, in the sense that we regard all space-time frameworks in which we can locate natural phenomena as on the same footing, yet we admit that space—the whole group of possible spaces—may have some absolute properties. It may, for instance, be homaloidal or non-homaloidal. Whatever the co-ordinates, space near attracting matter is non-homaloidal, space at an infinite distance from matter is homaloidal. You cannot use the same co-ordinates for describing both kinds of space, any more than you can use rectangular co-ordinates on the surface of a sphere; that is, in fact, the geometrical interpretation of the difference. Homaloidal space-time may be regarded as a four-dimensional plane drawn in a continuum of five dimensions; whereas non-homaloidal space-time must be regarded as a curved surface in five dimensions.* These considerations apply, of course, to *measured* space; we can always throw the blame on our measuring rods, and apply theoretical corrections to

* We shall see (§44) that in a region, not containing matter, but traversed by a gravitational field due to matter, the Gaussian or total curvature is zero; but such a space-time does not correspond to a plane in five dimensions, or to any surface which can be developed into a plane. The space-time in a gravitational field has an essential curvature in the ordinary sense, although it happens that the particular invariant technically called "the curvature" vanishes. In three-dimensional space a surface with zero Gaussian curvature can always be developed into a plane; but this is not true for space of higher dimensions, so that the three-dimensional analogy is liable to lead to misunderstanding.

our measures so as to make them agree with any kind of space we please.

It is not necessary, and indeed it is not possible, to draw a sharp distinction between the portions of the g's arising from the choice of co-ordinates and the portions arising from the gravitation of matter. We have seen that, when there is no field of force, ds^2 has the form (15·1), so that the g's have the values,

$$\begin{array}{cccc} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{array} \qquad (16{\cdot}3)$$

These values then express that there is no field of force, and in the absence of a gravitational field produced by matter it is possible to take our co-ordinates (Galilean co-ordinates) so that the values (16·3) hold everywhere. We naturally regard such co-ordinates as fundamental; and, if we choose any other co-ordinates, the deviations of the g's from this peculiarly simple set of values are regarded as due to the distortion of the space-time chosen. But by § 14, when gravitating matter is in the neighbourhood, there is no possibility of choosing co-ordinates, so that the values (16·3) hold everywhere, and there is no criterion for selecting any one of the possible systems of co-ordinates as more fundamental than the others.*

Accordingly we shall henceforth apply the term "gravitational field" to the whole field of force given by the g's, whatever its origin. In the particular case when no part of it is due to the gravitation of matter, we shall say there is no *permanent* gravitational field.

Just as Galilean co-ordinates are defined by the values (16·3) of the g's, so any other co-ordinates must be defined analytically by specifying the g's as functions of x_1, x_2, x_3, x_4, or—what comes to the same thing—by giving the expression for ds^2. For example, if in two dimensions $ds^2 = dx_1^2 + x_1^2 dx_2^2$, the co-ordinates are recognised as plane polar co-ordinates with

* Thus if we say "take rectangular axes with the sun as origin," the statement is ambiguous. Unaccelerated rectangular axes imply that ds^2 is of the form (15·1)—no other means of defining them having yet been given. Owing to the sun's gravitation there is no system of co-ordinates for which this is true, and several different systems present rival claims to be regarded as the best approximation possible. The difficulty does not arise if we only have to consider an infinitesimal region of space; in that case the co-ordinates (giving "natural measure") are defined without ambiguity.

$x_1=r$, $x_2=\theta$; if $ds^2=dx_1^2+\cos^2x_1dx_2^2$, the co-ordinates are latitude (x_1) and longitude (x_2) on a sphere. We might take for the 10 g's perfectly arbitrary functions of x_1, x_2, x_3, x_4, and so obtain a ten-fold infinity of mathematically conceivable systems of co-ordinates. But this would include many systems of co-ordinates which describe kinds of space-time not occurring in Nature. In any particular problem our choice is restricted to a four-fold infinity, viz., if x_1, x_2, x_3, x_4 is a possible system, then four arbitrary functions of x_1, x_2, x_3, x_4 will form a possible system. In some other problem there will be an entirely different group of possible systems; the space-times in the two problems have thus certain absolute properties which are irreconcilable, and we interpret this physically by saying that the permanent gravitational field is different in the two cases. Further, taking all possible distributions of permanent gravitational field which can occur in space (in the neighbourhood of, but not containing, matter), we do not exhaust the conceivable variety of functions expressing the g's. There is a general limitation on the g's—imposed, not by mathematics, but by Nature—which is expressed by the differential equations of the law of gravitation which we are about to seek. The law of gravitation, in fact, expresses certain absolute properties common to all the measured space-times that can under any conditions occur in Nature.

The law of gravitation, or general relation connecting the g's, must hold for all observed values of the g's. Since the g's define the system of co-ordinates used, this means that the relation must hold for all possible systems of co-ordinates. If new co-ordinates are chosen, we find new values of the g's as in (15·5); and the differential equations between the new g's and new co-ordinates must be the same as between the old g's and old co-ordinates. In mathematical language the equations must be covariant.

There is a resemblance between this statement and the statement of § 12 which is somewhat deceptive. We there found that observable events have no reference to any particular system of co-ordinates, and therefore all laws of nature can be expressed in a form independent of the co-ordinates. But this alone does not allow us to deduce the covariance of the equations satisfied by the gravitation-potentials. Without the principle of equivalence we could no doubt define the field by certain potentials φ_1, φ_2, φ_3, ... which satisfy differential equations independent of the choice of co-ordinates. But that

conveys no information of value, unless we are told how to find φ_1', φ_2', ... in the co-ordinates x_1', x_2', x_3', x_4' from the values φ_1, φ_2, ... in the co-ordinates x_1, x_2, x_3, x_4. The statement in § 12 tells us nothing about that. It is the principle of equivalence which, by identifying the potentials with the g's for which the method of transformation is known, supplies the missing link.

17. The Newtonian law of gravitation, $\nabla^2 g_{44}=0$, does not *fulfil the condition of covariance nor does any modification of it, which immediately suggests itself. We have, therefore, to* seek a new law guided by the condition that it must be expressed by a covariant set of equations between the g's. It will be found in Chapter IV. that the choice is so restricted as to leave little doubt as to what the new law must be.

If we write the required equations in the form

$$T_1=0, \quad T_2=0, \quad T_3=0, \quad \&c.,$$

the left-hand sides, T_1, T_2, T_3, may be regarded as components of a kind of generalised vector, only the number of components is not, as in a vector, restricted to 4.

The covariance of the equations means that, if all the components vanish in one system of co-ordinates, they must vanish in all systems. To secure this, T_1, T_2, ... must obey a linear law of transformation ; thus

$$\begin{aligned} T_1' &= \lambda_1 T_1 + \lambda_2 T_2 + \lambda_3 T_3 + \;\ldots\ldots \qquad (17\cdot 1)\\ T_2' &= \mu_1 T_1 + \mu_2 T_2 + \mu_3 T_3 + \;\ldots \end{aligned}$$

where the coefficients are functions of the co-ordinates depending on the transformation. Generalised vectors of this kind are called tensors ; and it will be necessary for us to study their properties in the next chapter, in order to select the one which can represent the new law of gravitation.

We see that if an equation is known to be a tensor-equation, it is sufficient to prove it for one particular system of co-ordinates ; it will then automatically hold in any other system obtainable by a mathematical transformation.

The more general purpose of the tensor theory is this : If we are given a set of equations expressing some physical law in the usual co-ordinates, we may be able to recognise these as the degenerate form for Galilean co-ordinates of some tensor equation. Expressed in tensor form, these equations will then hold for all systems of co-ordinates that can be derived by a mathematical transformation. Subject to the

limitations of § 14, they will also hold for the gravitational field, although the co-ordinates in that case cannot be obtained by a mathematical transformation. The intermediate step is of no great interest, since the mere transformation of co-ordinates leads to nothing new. But by this mode of approach we obtain the corresponding equations as modified by the action of a gravitational field. This is a very powerful method of investigation.

18. I anticipate that some readers will find the next two chapters difficult, and I therefore place here, out of order, a brief account of the field of a particle according to the new law of gravitation; but I doubt if there is any royal road to relativity, and it is scarcely possible to make serious progress except by analytical methods.

We shall find that when a heavy particle is at rest at the origin the expression for the line-element in plane polar co-ordinates is

$$ds^2 = -\gamma^{-1}dr^2 - r^2 d\theta^2 + \gamma dt^2 \quad . \quad . \quad . \quad . \quad (18\cdot1)$$

where $\gamma = 1 - 2m/r$, and m is the mass of the particle, the constant of gravitation and the velocity of light being unity. For the sun, $m = 1\cdot47$ kilometres*, so that γ generally differs from unity by a very small quantity.

If $\Omega = m/r$ = Newtonian potential at the point considered, we have

$$g_{44} = \gamma = 1 - 2\Omega \quad . \quad . \quad . \quad . \quad . \quad (18\cdot2)$$

just as in the case of the centrifugal force (16·2). The Newtonian attraction is therefore a consequence of the coefficient of dt^2.

The general meaning of (18·1) is that our measures will not fit together in Euclidean space. Measuring in the direction r we have

$$ids = \gamma^{-\frac{1}{2}}dr,$$

that is to say we must correct the measured length ids in the radial direction, multiplying it by $\gamma^{\frac{1}{2}}$ in order to obtain a length dr which will fit into Euclidean space. Or we may say that our measuring rod contracts when placed radially; transverse measures require no correction. Similarly the measured time must be multiplied by $\gamma^{-\frac{1}{2}}$, *i.e.*, our clocks run slow.

* This can be verified roughly as follows:—For a circular orbit $m/r^2 = v^2/r$, the constant of gravitation being unity. Applying this to the earth, v = earth's orbital velocity = 30 km. per sec. = 10^{-4} × velocity of light. Hence in our units $m = 10^{-8}r$; and r, the radius of the earth's orbit, is $1\cdot5 \times 10^8$ km.

But there is more than one way of correcting the measures to fit Euclidean space, so that we are not really justified in making precise statements as to the behaviour of our clocks and measuring rods. It is better not to discuss their defects, but to accept the measures and examine the properties of the corresponding non-Euclidean space and time.

If we draw a circle with a heavy particle near the centre, the ratio of the measured circumference to the measured diameter will be a little less than π, owing to the factor $\gamma^{-\frac{1}{2}}$ affecting radial measures. It is thus like a circle drawn on a sphere, for which the circumference is less than π times the diameter if we measure along the surface of the sphere. We may imagine space pervaded by a gravitational field to have a curvature in some purely mathematical fifth dimension.

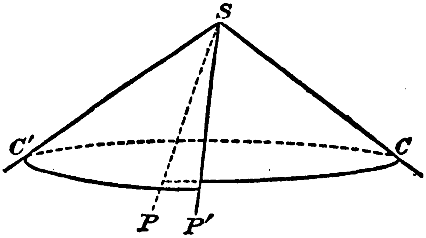

FIG. 3.

If we draw the elliptic orbit of a planet, slit it along a radius and try to fold it round our curved space there will evidently be some overlap. For example, take a cone with the sun as apex as roughly representing the curved space. Starting with the radius vector SP, the Euclidean space will fold completely round the cone and overlap to the extent PSP'. Thus the corresponding radius advances through an angle PSP' each revolution (Fig. 3). This shows one reason for the advance of perihelion of a planet, which is one of the most important effects predicted by the new theory; but it is not the whole explanation.

The reader may not unnaturally suspect that there is an admixture of metaphysics in a theory which thus reduces the gravitational field to a modification of the metrical properties of space and time. This suspicion, however, is a complete misapprehension, due to the confusion of space, as we have defined it, with some transcendental and philosophical space.

There is nothing metaphysical in the statement that under certain circumstances the measured circumference of a circle is less than π times the measured diameter; it is purely a matter for experiment. We have simply been studying the way in which physical measures of length and time fit together —just as Maxwell's equations describe how electrical and magnetic forces fit together. The trouble is that we have inherited a preconceived idea of the way in which measures, if "true," ought to fit. But the relativity standpoint is that we do not know, and do not care, whether the measures under discussion are "true" or not; and we certainly ought not to be accused of metaphysical speculation, since we confine ourselves to the geometry of measures which are strictly practical, if not strictly practicable. It is desirable to insist that we do not attribute any *causative* properties to these distortions of measured space and time. To hold that a property of our measuring-rods is the cause of gravitation would be as absurd as to hold that the fall of the barometer is the cause of the storm.

CHAPTER III.

THE THEORY OF TENSORS.

19. We consider transformations from one system of coordinates x_1, x_2, x_3, x_4 to another system x_1', x_2', x_3', x_4'.

(*a*) *Notation.*

The formula (15·3) for ds^2 may be written

$$ds^2 = \sum_{\mu=1}^{4} \sum_{\nu=1}^{4} g_{\mu\nu} dx_\mu dx_\nu \quad (g_{\mu\nu} = g_{\nu\mu}) \quad . \quad . \quad . \quad (19{\cdot}11)$$

In the following work we shall omit the signs of summation, adopting the convention that, whenever a literal suffix appears twice in a term, the term is to be summed for values of the suffix 1, 2, 3, 4. If a suffix appears once only, no summation is indicated. Thus we shall write (19·11)

$$ds^2 = g_{\mu\nu} dx_\mu dx_\nu \quad . \quad . \quad . \quad . \quad . \quad . \quad (19{\cdot}12)$$

In rare cases it may be necessary to write a term containing a suffix twice which is not to be summed; these cases will always be specially indicated. In general, however, this convention anticipates our desires, and actually gives a kind of momentum in the right direction to the analysis.

As a rule of manipulation it may be noticed that any suffix appearing twice is a dummy, and can be changed freely to any other suffix not occurring in the same term.

(*b*) *Covariant and Contravariant Vectors.*

The vector (dx_1, dx_2, dx_3, dx_4) is transformed according to the equations

§250

$$dx_1' = \frac{\partial x_1'}{\partial x_1} dx_1 + \frac{\partial x_1'}{\partial x_2} dx_2 + \frac{\partial x_1'}{\partial x_3} dx_3 + \frac{\partial x_1'}{\partial x_4} dx_4, \text{ \&c.}$$

or, with our convention as to notation

$$dx_\mu' = \frac{\partial x_\mu'}{\partial x_\sigma} dx_\sigma.$$

Any vector transformed according to this law is called a *contravariant* vector; its character is denoted by the notation A^μ ($\mu=1, 2, 3, 4$). The law may be written 82 50

$$A'^\mu=\frac{\partial x'_\mu}{\partial x_\sigma}A^\sigma \quad . \quad . \quad . \quad . \quad . \quad . \quad (19\cdot21)$$

where, as already explained, summation is indicated by the double appearance of the dummy σ.

If φ is a scalar (*i.e.*, invariant) function of position the vector $\left(\frac{\partial\varphi}{\partial x_1}, \frac{\partial\varphi}{\partial x_2}, \frac{\partial\varphi}{\partial x_3}, \frac{\partial\varphi}{\partial x_4}\right)$ is transformed according to the law

$$\frac{\partial\varphi}{\partial x'_1}=\frac{\partial x_1}{\partial x'_1}\frac{\partial\varphi}{\partial x_1}+\frac{\partial x_2}{\partial x'_1}\frac{\partial\varphi}{\partial x_2}+\frac{\partial x_3}{\partial x'_1}\frac{\partial\varphi}{\partial x_3}+\frac{\partial x_4}{\partial x'_1}\frac{\partial\varphi}{\partial x_4}$$

A vector transformed according to this law is called a *covariant* vector, denoted by A_μ. The law may be written

$$A'_\mu=\frac{\partial x_\sigma}{\partial x'_\mu}A_\sigma \quad . \quad . \quad . \quad . \quad . \quad . \quad (19\cdot22)$$

A covariant vector is not necessarily the gradient of a scalar.

The customary geometrical conception of a vector does not reveal the distinction between the two classes of contravariant and covariant vectors. We usually represent any directed quantity by a straight line, which should strictly correspond only to a contravariant vector. The other class of directed quantities is more properly represented by the reciprocal of a straight line; but in elementary applications, when we are thinking in terms of rectangular co-ordinates, there is no need to make this distinction. Consider, however, a fluid with a velocity potential. With rectangular co-ordinates the velocity is equal to the gradient of the velocity potential. Both these are directed quantities, *i.e.*, vectors, and the vector relation extends to their rectangular components; thus—

$$\frac{dx}{dt}=\frac{\partial\varphi}{\partial x}, \quad \frac{dy}{dt}=\frac{\partial\varphi}{\partial y}, \quad \frac{dz}{dt}=\frac{\partial\varphi}{\partial z}.$$

But if we use oblique axes or curvilinear co-ordinates, the relation no longer holds. *E.g.*, it is not true that in polar co-ordinates $d\theta/dt=\partial\varphi/\partial\theta$; the actual relation is $rd\theta/dt=\partial\varphi/rd\theta$. This is because the two vectors are of opposite natures, the first being contravariant and the second covariant. If they had been of the same nature the relation must have held for all systems of co-ordinates. Clearly, since in our work we consider all systems of co-ordinates as on the same footing, we have to distinguish carefully between the two types. We

realise at once that the equation $dx_\mu/dt = \partial\phi/\partial x_\mu$, being an equation between vectors of opposite kinds, is impossible as a general equation for all systems of co-ordinates, *i.e.*, it is not a covariant equation.

(*c*) *Tensors of Higher Rank.*

We can denote by $A_{\mu\nu}$ a quantity having 16 components, obtained by giving different numerical values to μ and ν. Similarly, $A_{\mu\nu\sigma}$ has 64 components. By a generalisation of (19·21) and (19·22) we classify quantities of this kind according to their transformation laws, viz.,

$$\text{Covariant tensors} \qquad A'_{\mu\nu} = \frac{\partial x_\sigma}{\partial x'_\mu}\frac{\partial x_\tau}{\partial x'_\nu}A_{\sigma\tau} \quad . \quad . \quad . \quad . \quad (19·31)$$

$$\text{Contravariant tensors} \qquad A'^{\mu\nu} = \frac{\partial x'_\mu}{\partial x_\sigma}\frac{\partial x'_\nu}{\partial x_\tau}A^{\sigma\tau} \quad . \quad . \quad . \quad . \quad (19·32)$$

$$\text{Mixed tensors} \qquad A'^{\nu}_{\mu} = \frac{\partial x_\sigma}{\partial x'_\mu}\frac{\partial x'_\nu}{\partial x_\tau}A^{\tau}_{\sigma} \quad . \quad . \quad . \quad . \quad (19·33)$$

and similarly for tensors of the third and higher rank. These equations of transformation are linear, so that the conditions of § 17 are satisfied. Also it is not difficult to see that there can be no other linear types of transformation-laws having the necessary transitive property. For example, consider a vector A_σ. Introducing a third set of co-ordinates x''_λ, we have

$$A''_\lambda = \frac{\partial x'_\mu}{\partial x''_\lambda}A'_\mu \quad \text{and} \quad A'_\mu = \frac{\partial x_\sigma}{\partial x'_\mu}A_\sigma.$$

$$\text{But} \qquad \frac{\partial x_\sigma}{\partial x'_\mu}\frac{\partial x'_\mu}{\partial x''_\lambda} = \frac{\partial x_\sigma}{\partial x''_\lambda}; \quad \text{hence } A''_\lambda = \frac{\partial x_\sigma}{\partial x''_\lambda}A_\sigma,$$

showing that the result is the same whether the transformation is performed in two steps or directly. Other suggested types of transformation law have not this necessary property. Thus all possible types of tensors are included.

Evidently the sum of two tensors of the same character is a tensor.

The product of two tensors is a tensor, and its character is the sum of the characters of the component tensors. For example, consider the product $A_{\mu\nu}B^{\rho}_{\lambda}$, we have by (19·31) and (19·33)

$$A'_{\mu\nu} = \frac{\partial x_\alpha}{\partial x'_\mu}\frac{\partial x_\beta}{\partial x'_\nu}A_{\alpha\beta}, \qquad B'^{\rho}_{\lambda} = \frac{\partial x_\gamma}{\partial x'_\lambda}\frac{\partial x'_\rho}{\partial x_\delta}B^{\delta}_{\gamma}.$$

$$\text{Hence} \qquad (A'_{\mu\nu}B'^{\rho}_{\lambda}) = \frac{\partial x_\alpha}{\partial x'_\mu}\frac{\partial x_\beta}{\partial x'_\nu}\frac{\partial x_\gamma}{\partial x'_\lambda}\frac{\partial x'_\rho}{\partial x_\delta}(A_{\alpha\beta}B^{\delta}_{\gamma}) \quad . \quad . \quad (19·34)$$

showing that the law of transformation is that of a tensor of the fourth rank having the character denoted by $C^{\rho}_{\mu\nu\lambda}$.

The product of two vectors is a tensor of the second rank, but a tensor of the second rank is not necessarily the product of two vectors.

A familiar example of a tensor of the second rank is afforded by the stresses in a solid or viscous fluid. The component of stress denoted by p_{xy} represents the traction in the y-direction exerted across an interface perpendicular to the x-direction. Each component involves a specification of two directions.

(*d*) *Inner Multiplication.*

If we multiply A_{μ} by B^{μ}, the repetition of the suffix involves summation of the resulting products. The result is called the *inner product* in contrast to the ordinary or *outer product* $A_{\mu}B^{\nu}$. The notation at once shows whether the product is inner or outer in any formula.

From a mixed tensor such as $A^{\tau}_{\mu\nu\sigma}$ we can form a "*contracted*" tensor $A^{\sigma}_{\mu\nu\sigma}$, which is of the second rank with suffixes μ and ν (since σ is now a dummy suffix). To show that it is a tensor we have as in (19·34)

$$A^{\sigma}_{\mu\nu\sigma}=\frac{\partial x_{\alpha}}{\partial x'_{\mu}}\frac{\partial x_{\beta}}{\partial x'_{\nu}}\frac{\partial x_{\gamma}}{\partial x'_{\sigma}}\frac{\partial x'_{\sigma}}{\partial x_{\delta}}A^{\delta}_{\alpha\beta\gamma} \quad . \quad . \quad . \quad (19·41)$$

But $\dfrac{\partial x_{\gamma}}{\partial x'_{\sigma}}\dfrac{\partial x'_{\sigma}}{\partial x_{\delta}}=\dfrac{\partial x_{\gamma}}{\partial x_{\delta}}=0$ or 1, according as $\gamma \neq \delta$ or $\gamma=\delta$.

Hence $\dfrac{\partial x_{\gamma}}{\partial x'_{\sigma}}\dfrac{\partial x'_{\sigma}}{\partial x_{\delta}}A^{\delta}_{\alpha\beta\gamma}=0+0+0+A^{\gamma}_{\alpha\beta\gamma}$.

Substituting in (19·41) we see that $A^{\sigma}_{\mu\nu\sigma}$ follows the law of transformation (19·31) and is therefore a covariant tensor.

An expression such as $A^{\tau}_{\mu\sigma\sigma}$ is not a tensor, and no interest attaches to it.

By a similar argument we see that A^{μ}_{μ}, $A^{\mu\nu}_{\mu\nu}$ are invariant, and consequently $A_{\mu}B^{\mu}$ is an invariant. An invariant, or scalar, corresponds to a tensor of zero rank.

(*e*) *Criterion for the Tensor Character.*

To prove that a given quantity is a tensor, we either find directly its equations of transformation, or we express it as the sum or product of other tensors, or, under certain restrictions, as the quotient of two tensors according to the following theorem: A quantity, which on inner multiplication by *any* covariant (alternatively, by *any* contravariant) vector always gives a tensor, is itself a tensor.

21. *Auxiliary Formula for the Second Derivatives.*

We introduce certain quantities known as Christoffel's 3-index symbols, viz.,

$$[\mu\nu, \lambda] = \tfrac{1}{2}\left(\frac{\partial g_{\mu\lambda}}{\partial x_\nu} + \frac{\partial g_{\nu\lambda}}{\partial x_\mu} - \frac{\partial g_{\mu\nu}}{\partial x_\lambda}\right) \quad . \quad . \quad . \quad (21 \cdot 11)$$

$$\{\mu\nu, \lambda\} = \tfrac{1}{2} g^{\lambda\alpha}\left(\frac{\partial g_{\mu\alpha}}{\partial x_\nu} + \frac{\partial g_{\nu\alpha}}{\partial x_\mu} - \frac{\partial g_{\mu\nu}}{\partial x_\alpha}\right) \quad . \quad (21 \cdot 12)$$

We have $\quad \{\mu\nu, \lambda\} = g^{\lambda\alpha}\,[\mu\nu, \alpha] \quad . \quad . \quad . \quad . \quad (21 \cdot 13)$

and the reciprocal relation follows by (20·1)

$$[\mu\nu, \lambda] = g_{\lambda\alpha}\,\{\mu\nu, \alpha\} \quad . \quad . \quad . \quad . \quad (21 \cdot 14)$$

Since $g_{\mu\nu}$ is a covariant tensor

$$g'_{\mu\nu} = \frac{\partial x_\alpha}{\partial x'_\mu}\frac{\partial x_\beta}{\partial x'_\nu} g_{\alpha\beta}.$$

Hence

$$\frac{\partial g'_{\mu\nu}}{\partial x'_\lambda} = g_{\alpha\beta}\left\{\frac{\partial^2 x_\alpha}{\partial x'_\mu \partial x'_\lambda}\frac{\partial x_\beta}{\partial x'_\nu} + \frac{\partial^2 x_\alpha}{\partial x'_\nu \partial x'_\lambda}\frac{\partial x_\beta}{\partial x'_\mu}\right\} + \frac{\partial x_\alpha}{\partial x'_\mu}\frac{\partial x_\beta}{\partial x'_\nu}\frac{\partial x_\gamma}{\partial x'_\lambda}\frac{\partial g_{\alpha\beta}}{\partial x_\gamma} \quad . \quad (21 \cdot 15)$$

In the second term in the bracket we have interchanged α and β, which is legitimate since they are dummies; in the last term we have used

$$\frac{\partial}{\partial x'_\lambda} = \frac{\partial x_\gamma}{\partial x'_\lambda}\frac{\partial}{\partial x_\gamma}.$$

Similarly,

$$\frac{\partial g'_{\nu\lambda}}{\partial x'_\mu} = g_{\alpha\beta}\left\{\frac{\partial^2 x_\alpha}{\partial x'_\nu \partial x'_\mu}\frac{\partial x_\beta}{\partial x'_\lambda} + \frac{\partial^2 x_\alpha}{\partial x'_\lambda \partial x'_\mu}\frac{\partial x_\beta}{\partial x'_\nu}\right\} + \frac{\partial x_\alpha}{\partial x'_\mu}\frac{\partial x_\beta}{\partial x'_\nu}\frac{\partial x_\gamma}{\partial x'_\lambda}\frac{\partial g_{\beta\gamma}}{\partial x_\alpha} \quad . \quad (21 \cdot 16)$$

$$\frac{\partial g'_{\mu\lambda}}{\partial x'_\nu} = g_{\alpha\beta}\left\{\frac{\partial^2 x_\alpha}{\partial x'_\mu \partial x'_\nu}\frac{\partial x_\beta}{\partial x'_\lambda} + \frac{\partial^2 x_\alpha}{\partial x'_\lambda \partial x'_\nu}\frac{\partial x_\beta}{\partial x'_\mu}\right\} + \frac{\partial x_\alpha}{\partial x'_\mu}\frac{\partial x_\beta}{\partial x'_\nu}\frac{\partial x_\gamma}{\partial x'_\lambda}\frac{\partial g_{\alpha\gamma}}{\partial x_\beta} \quad . \quad (21 \cdot 17)$$

where in the last term we have made some interchanges of the dummy suffixes α, β, γ.

Adding these two equations and subtracting (21·15) we have by (21·11)

$$[\mu\nu, \lambda]' = g_{\alpha\beta}\frac{\partial^2 x_\alpha}{\partial x'_\mu \partial x'_\nu}\frac{\partial x_\beta}{\partial x'_\lambda} + \frac{\partial x_\alpha}{\partial x'_\mu}\frac{\partial x_\beta}{\partial x'_\nu}\frac{\partial x_\gamma}{\partial x'_\lambda}[\alpha\beta, \gamma] \quad . \quad (21 \cdot 18)$$

Multiply through by $g'^{\lambda\rho}\dfrac{\partial x_\epsilon}{\partial x'_\rho}$, we have

$$\{\mu\nu, \rho\}'\frac{\partial x_\epsilon}{\partial x'_\rho} = \left(g'^{\lambda\rho}\frac{\partial x_\beta}{\partial x'_\lambda}\frac{\partial x_\epsilon}{\partial x'_\rho}\right) g_{\alpha\beta}\frac{\partial^2 x_\alpha}{\partial x'_\mu\, \partial x'_\nu} + \left(g'^{\lambda\rho}\frac{\partial x_\gamma}{\partial x'_\lambda}\frac{\partial x_\epsilon}{\partial x'_\rho}\right)\frac{\partial x_\alpha}{\partial x'_\mu}\frac{\partial x_\beta}{\partial x'_\nu}[\alpha\beta, \gamma]$$

$$= g^{\beta\epsilon} g_{\alpha\beta}\frac{\partial^2 x_\alpha}{\partial x'_\mu\, \partial x'_\nu} + \frac{\partial x_\alpha}{\partial x'_\mu}\frac{\partial x_\beta}{\partial x'_\nu} g^{\gamma\epsilon}\,[\alpha\beta, \gamma] \text{ by } (19 \cdot 32)$$

$$= \frac{\partial^2 x_\epsilon}{\partial x'_\mu\, \partial x'_\nu} + \frac{\partial x_\alpha}{\partial x'_\mu}\frac{\partial x_\beta}{\partial x'_\nu}\{\alpha\beta, \epsilon\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (21 \cdot 2)$$

using (20·1) and (21·13).

This somewhat complicated formula for $\partial^2 x_\epsilon / \partial x'_\mu \partial x'_\nu$ in terms of the first derivatives is needed for the developments in the next paragraph.

22. Covariant Differentiation.

If we differentiate a scalar quantity we obtain a tensor (a covariant vector); but if we differentiate a tensor of the first or higher rank the result is not a tensor. We can, however, obtain a tensor which plays the part of a derivative by a more general process. The process is particularly useful in generalising results which have been obtained in Galilean co-ordinates, since the simple derivative is the degenerate form for Galilean co-ordinates of the covariant derivative here considered.

If A_μ is a covariant vector, then by (19·22)

$$A'_\mu = \frac{\partial x_\sigma}{\partial x'_\mu} A_\sigma.$$

Whence, differentiating,

$$\frac{\partial A'_\mu}{\partial x'_\nu} = \frac{\partial x_\sigma}{\partial x'_\mu}\frac{\partial x_\tau}{\partial x'_\nu}\frac{\partial A_\sigma}{\partial x_\tau} + \frac{\partial^2 x_\sigma}{\partial x'_\mu \partial x'_\nu} A_\sigma.$$

Substitute for $\partial^2 x_\sigma / \partial x'_\mu \partial x'_\nu$ by (21·2); we have

$$\frac{\partial A'_\mu}{\partial x'_\nu} - \{\mu\nu, \rho\}' A_\sigma \frac{\partial x_\sigma}{\partial x'_\rho} = \frac{\partial x_\sigma}{\partial x'_\mu}\frac{\partial x_\tau}{\partial x'_\nu}\frac{\partial A_\sigma}{\partial x_\tau} - \frac{\partial x_\alpha}{\partial x'_\mu}\frac{\partial x_\beta}{\partial x'_\nu}\{\alpha\beta, \sigma\} A_\sigma \quad (22·1)$$

But $A_\sigma \dfrac{\partial x_\sigma}{\partial x'_\rho} = A'_\rho$ by (19·22); and in the last term the dummies α, β, σ may be replaced by σ, τ, ρ. Hence if we write

$$A_{\mu\nu} = \frac{\partial A_\mu}{\partial x_\nu} - \{\mu\nu, \rho\} A_\rho \quad . \quad . \quad . \quad . \quad (22·2)$$

we have
$$A'_{\mu\nu} = \frac{\partial x_\sigma}{\partial x'_\mu}\frac{\partial x_\tau}{\partial x'_\nu} A_{\sigma\tau},$$

showing that $A_{\mu\nu}$ is a tensor. This is called the *covariant derivative* of A_μ.

If A_λ, B_μ are covariant vectors, $A_{\lambda\nu}$, $B_{\mu\nu}$ their covariant derivatives, then

$$A_{\lambda\nu} B + A_\lambda B_{\mu\nu}$$

is the sum of two tensors, and is therefore a tensor. Substituting from (22·2) this tensor becomes

$$\frac{\partial (A_\lambda B_\mu)}{\partial x_\nu} - \{\lambda\nu, \epsilon\} A_\epsilon B_\mu - \{\mu\nu, \epsilon\} A_\lambda B_\epsilon \quad . \quad (22·3)$$

which is called the derivative of the tensor $A_\lambda B_\mu$. It is not difficult to show that any tensor of the second rank can be expressed as the sum of products of four pairs of vectors, and hence (22·3) can be generalised, giving for the covariant derivative of $A_{\lambda\mu}$

$$A_{\lambda\mu\nu}=\frac{\partial A_{\lambda\mu}}{\partial x_\nu}-\{\lambda\nu,\,\varepsilon\}A_{\epsilon\mu}-\{\mu\nu,\,\varepsilon\}A_{\lambda\epsilon} \quad . \quad . \quad (22·4)$$

In a somewhat similar manner formulæ for the covariant derivatives of contravariant and mixed tensors can be obtained, viz.,

$$A^\mu_\nu=\frac{\partial A^\mu}{\partial x_\nu}+\{\nu\varepsilon,\,\mu\}A^\epsilon \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (22·5)$$

$$A^{\lambda\mu}_\nu=\frac{\partial A^{\lambda\mu}}{\partial x_\nu}+\{\nu\varepsilon,\,\lambda\}A^{\epsilon\mu}+\{\nu\varepsilon,\,\mu\}A^{\lambda\epsilon} \quad . \quad . \quad (22·6)$$

$$A^\mu_{\lambda\nu}=\frac{\partial A^\mu_\lambda}{\partial x_\nu}-\{\nu\lambda,\,\varepsilon\}A^\mu_\epsilon+\{\nu\varepsilon,\,\mu\}A^\epsilon_\lambda \quad . \quad . \quad . \quad (22·7)$$

The unsymmetrical behaviour of covariant and contravariant indices in these formulæ should be noticed. In all cases differentiation adds one unit of covariant character.

When the g's have Galilean values (or, more generally, are constants) the Christoffel symbols vanish, and these derivatives reduce in all cases to the ordinary differential coefficients.

23. *The Riemann-Christoffel Tensor.*

Let us form the second covariant derivative of the vector A_μ, that is to say in formula (22·4) we give the tensor $A_{\lambda\mu}$ the value (22·2).

$$A_{\mu\nu\sigma}=\frac{\partial}{\partial x_\sigma}\left\{\frac{\partial A_\mu}{\partial x_\nu}-\{\mu\nu,\,\rho\}A_\rho\right\}-\{\mu\sigma,\,\varepsilon\}\left\{\frac{\partial A_\epsilon}{\partial x_\nu}-\{\varepsilon\nu,\,\rho\}A_\rho\right\}$$
$$-\{\nu\sigma,\,\varepsilon\}\left\{\frac{\partial A_\mu}{\partial x_\epsilon}-\{\mu\varepsilon,\rho\}A_\rho\right\}$$
$$=\frac{\partial^2 A_\mu}{\partial x_\sigma\,\partial x_\nu}-\{\mu\nu,\,\rho\}\frac{\partial A_\rho}{\partial x_\sigma}-\{\mu\sigma,\,\varepsilon\}\frac{\partial A_\epsilon}{\partial x_\nu}-\{\nu\sigma,\,\varepsilon\}\frac{\partial A_\mu}{\partial x_\epsilon}$$
$$+\{\nu\sigma,\,\varepsilon\}\{\mu\varepsilon,\,\rho\}A_\rho+\{\mu\sigma,\,\varepsilon\}\{\varepsilon\nu,\,\rho\}A_\rho-A_\rho\frac{\partial}{\partial x_\sigma}\{\mu\nu,\rho\}$$

The first five terms are unaltered by interchanging ν and σ, *i.e.*, by changing the order of differentiation. (We can write ε for ρ in the second term.) Hence

$$A_{\mu\nu\sigma}-A_{\mu\sigma\nu}=$$
$$\left[\{\mu\sigma,\,\varepsilon\}\{\varepsilon\nu,\,\rho\}-\{\mu\nu,\,\varepsilon\}\{\varepsilon\sigma,\,\rho\}+\frac{\partial}{\partial x_\nu}\{\mu\sigma,\,\rho\}-\frac{\partial}{\partial x_\sigma}\{\mu\nu,\,\rho\}\right]A_\rho$$

The left side is a tensor, and A_ρ is an arbitrary covariant vector; therefore, by § 19 (*e*) the quantity in the bracket is a tensor. This is called the Riemann-Christoffel tensor, and is denoted by

$$B^\rho_{\mu\nu\sigma}=\{\mu\sigma,\epsilon\}\{\epsilon\nu,\rho\}-\{\mu\nu,\epsilon\}\{\epsilon\sigma,\rho\}+\frac{\partial}{\partial x_\nu}\{\mu\sigma,\rho\}-\frac{\partial}{\partial x_\sigma}\{\mu\nu,\rho\} \quad (23)$$

24. *Conditions for Vanishing of the Riemann-Christoffel Tensor.*

From the foregoing definition the primary meaning of the vanishing of this tensor is that the order of differentiation is indifferent (as in the ordinary differentiation). But the tensor has an even more important property. It will be seen on inspection that it vanishes when the g's have their constant Galilean values.* But, since it is a tensor, it must also vanish in any other system of co-ordinates derivable by a mathematical transformation. Thus the equation

$$B^\rho_{\mu\nu\sigma}=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (24{\cdot}1)$$

is a necessary condition that with suitable choice of co-ordinates ds^2 can be reduced to the form

$$ds^2=-dx^2-dy^2-dz^2+dt^2. \quad . \quad . \quad . \quad (24{\cdot}2)$$

In other words it is a necessary condition for the absence of a permanent gravitational field.

It can be shown that the condition is also sufficient. Equation (24·1) contains 96 apparently different equations, since, owing to the antisymmetry in σ and ν, there are only 6 combinations of σ and ν to be combined with 16 combinations of μ and ρ. But these are not all independent, and the number can be reduced to 20, which can be shown to be the number of conditions required for the transformation to the form (24·2) to be possible.

The reduction is effected by writing

$$(\mu\tau\sigma\nu)=g_{\tau\rho}B^\rho_{\mu\nu\sigma},$$

so that $\qquad B^\rho_{\mu\nu\sigma}=g^{\lambda\rho}(\mu\lambda\sigma\nu) \qquad$ by (20·1)

Equation (24·1) is thus equivalent to

$$(\mu\tau\sigma\nu)=0,$$

and *vice versa*.

* The Christoffel symbols vanish when the g's are constants.

On working out the value of $(\mu\tau\sigma\nu)$ it is seen by inspection that the following additional relations exist :—

$$(\mu\tau\sigma\nu) \equiv -(\tau\mu\sigma\nu) \equiv (\nu\sigma\tau\mu) \equiv (\sigma\nu\mu\tau),$$

$$(\mu\tau\sigma\nu) + (\mu\sigma\nu\tau) + (\mu\nu\tau\sigma) \equiv 0,$$

which reduce the number of independent conditions to 20.

25. To sum up what has been accomplished in this chapter, we have discussed the theory of tensors—expressions which have the property that a linear relation between tensors of the same character will hold in all systems of co-ordinates if it holds in one system. We have shown that the tensor-property can be established either by determining the law of transformation, or exhibiting the quantity as a sum or product of other tensors, or, under certain restrictions, as the quotient of tensors. We have found formulæ for tensors which play the part of derivatives. Finally, we have found the necessary and sufficient relation between the $g_{\mu\nu}$, which must be satisfied in all systems of co-ordinates, when there is no permanent gravitational field.

This last result is an important step towards obtaining the law of gravitation. Any set of values of the g's which satisfy (24·1) will correspond to a *possible* set of co-ordinates which can be used for describing space not containing a permanent gravitational field. Hence if (24·1) is satisfied the g's are such as can occur in Nature, and are accordingly not inconsistent with the law of gravitation. The required equations of the law of gravitation must, therefore, include the vanishing of the Riemann-Christoffel tensor as a special case.

CHAPTER IV.

EINSTEIN'S LAW OF GRAVITATION.

26. We have seen in § 16 that the law of gravitation must be expressed as a set of differential equations satisfied by the g's. We have further found the equations (24·1) which are satisfied in the absence of (*i.e.*, at an infinite distance from) attracting matter. Clearly the general equations between the g's must be covariant equations automatically satisfied when (24·1) is satisfied; but they must be less stringent, so as to admit of permanent gravitational fields, which, we know, do not satisfy (24·1).

The simplest set of equations that suggests itself is

$$G_{\mu\nu} \equiv B^{\rho}_{\mu\nu\rho} = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (26{\cdot}1)$$

$G_{\mu\nu}$ being the contracted Riemann-Christoffel tensor, formed by setting $\sigma = \rho$ and summing. It is evidently satisfied when all components of the Riemann-Christoffel tensor vanish; and it is a less stringent condition.

The equations $G_{\mu\nu} = 0$ are taken by Einstein for the Law of Gravitation. Written in full they are, by (23)

$$-\frac{\partial}{\partial x_\rho}\{\mu\nu, \rho\} + \{\mu\rho, \epsilon\}\{\nu\epsilon, \rho\} + \frac{\partial}{\partial x_\nu}\{\mu\rho, \rho\} - \{\mu\nu, \epsilon\}\{\epsilon\rho, \rho\} = 0 \qquad (26{\cdot}2)$$

The last two terms can be simplified. We have

$$\{\mu\rho, \rho\} = \tfrac{1}{2}g^{\rho\epsilon}\left(\frac{\partial g_{\mu\epsilon}}{\partial x_\rho} + \frac{\partial g_{\rho\epsilon}}{\partial x_\mu} - \frac{\partial g_{\mu\rho}}{\partial x_\epsilon}\right)$$

$$= \tfrac{1}{2}g^{\rho\epsilon}\frac{\partial g_{\rho\epsilon}}{\partial x_\mu},$$

the other terms cancelling on summation.

Hence, since $g^{\rho\epsilon}g$ is the minor of the element $g_{\rho\epsilon}$ in the determinant g,

$$\{\mu\rho, \rho\} = \frac{1}{2g}\frac{\partial g}{\partial x_\mu} = \frac{\partial}{\partial x_\mu}\log\sqrt{-g}. \quad . \quad . \quad (26{\cdot}25)$$

Equation (26·2) thus becomes

$$G_{\mu\nu}=-\frac{\partial}{\partial x_\rho}\{\mu\nu,\rho\}+\{\mu\rho,\epsilon\}\{\nu\epsilon,\rho\}+\frac{\partial^2}{\partial x_\mu\partial x_\nu}\log\sqrt{-g}$$

$$-\{\mu\nu,\epsilon\}\frac{\partial}{\partial x_\epsilon}\log\sqrt{-g}=0 \quad . \quad (26{\cdot}3)$$

The equation is symmetrical in μ and ν, and therefore represents 10 different equations. Actually there exist four identical relations between these, so that the number of independent equations is reduced to six (*see* § 39).

The selection of this law of gravitation is not so arbitrary as it might appear. There is no other set of equations corresponding to a tensor of the second rank containing only first and second derivatives of the $g_{\mu\nu}$ and linear in the second derivatives. Moreover, there is no other way of building up a tensor of lower rank out of the components of $B^\rho_{\mu\nu\sigma}$.*

Having regard to the summations involved in (26·3) it will be seen that the application of the new law of gravitation must involve a considerable amount of calculation. There are first to be calculated 40 different Christoffel symbols, each of which is the sum of 12 terms. Then each of the 10 equations contains 25 terms—chiefly products or derivatives of the Christoffel symbols. Finally the partial differential equations have to be solved. It will probably be admitted that it is worth while to find out whether this suggested law of gravitation will agree with observation before resorting to something more complicated.

27. We are now in a position to define the Principle of Equivalence more precisely. The difference between a permanent gravitational field and an artificial one arising from a transformation of Galilean co-ordinates is that in the latter case (24·1) is satisfied, whereas in the former the less stringent condition (26·1) is satisfied. These equations determine the second differential coefficients of the $g_{\mu\nu}$, so that we can make the natural and artificial fields correspond as far as first differential coefficients, but not in the second differential coefficients. We shall therefore state the Principle of Equivalence as follows :—

* The tensor $B^\rho_{\rho\nu\sigma}$ vanishes identically. Other suggestions such as $g^{\mu\sigma}B^\rho_{\mu\nu\rho}$ merely give a set of equations equivalent to (26·1). The single equation $g^{\mu\nu}G_{\mu\nu}=0$ would obviously be insufficient to determine the gravitational field.

All laws, relating to phenomena in a geometrical field of force, *which depend on the g's and their first derivatives*, will also hold in a permanent gravitational field. Laws which depend on the second derivatives of the g's will not necessarily apply.

It must be remembered that we give no proof of this; it is merely an explicit statement of our assumptions. It would be quite consistent with the general idea of relativity if the true expression of such laws involved the Riemann-Christoffel tensor, which vanishes in the artificial field, and would have to be replaced before the equations were applied to the gravitational field. But if we were to admit that, the principle of equivalence would become absolutely useless.

THE GRAVITATIONAL FIELD OF A PARTICLE.

28. We have seen that the gravitational-potentials satisfy the equations (26·3)

$$G_{\sigma\tau} \equiv -\frac{\partial}{\partial x_\alpha}\{\sigma\tau, \alpha\} + \{\sigma\alpha, \beta\}\{\tau\beta, \alpha\} + \frac{\partial^2}{\partial x_\sigma \partial x_\tau}\log\sqrt{-g}$$
$$-\{\sigma\tau, \alpha\}\frac{\partial}{\partial x_\alpha}\log\sqrt{-g} = 0. \quad . \quad (28{\cdot}1)$$

We shall now find a solution of these equations corresponding to the field of a particle at rest at the origin of space-co-ordinates. We choose polar co-ordinates, viz.,

$$x_1 = r, \quad x_2 = \theta, \quad x_3 = \phi, \quad x_4 = t.$$

In making this statement we are departing somewhat from the standpoint of general relativity. Strictly speaking, we can only define a system of co-ordinates by the form of the corresponding expression for ds^2, that is by the gravitation-potentials. So that to specify the co-ordinates that are used involves solving the problem. Further, we have at present no knowledge of a particle of matter, except that it must be a point where the equations (28·1), which hold at points outside matter, break down; we can only distinguish a particle from other mathematically possible singularities, such as doublets, by the symmetry of the resulting field. Thus the logical course is to find a solution, and afterwards discuss what distribution of matter and what system of co-ordinates it represents. We shall, however, find it more profitable to accept the guidance of our current approximate ideas in order to arrive at the required solution inductively.

The line-element ds can be assumed to be of the form

$$ds^2 = -e^\lambda dr^2 - e^\mu(r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2) + e^\nu dt^2 \quad . \quad (28{\cdot}21)$$

where λ, μ, ν are functions of r only.

The omission of the product terms, $drd\theta$, $drd\phi$, $d\theta d\phi$, is justified by the symmetry of polar co-ordinates; the omission of $drdt$, $d\theta dt$, $d\phi dt$ involves the symmetry of a static field with respect to past and future time. If the latter products were present we should interpret the co-ordinates as changing with the time.

A further simplification can be made by writing $r^2 e^\mu = r'^2$ and adopting r' as our new co-ordinate (dropping the accent). The resulting change in dr^2 is absorbed by taking a new λ. Thus the coefficient e^μ is made to disappear and we have

$$ds^2 = -e^\lambda dr^2 - r^2 d\theta^2 - r^2 \sin^2\theta d\phi^2 + e^\nu dt^2. \quad . \quad (28{\cdot}22)$$

Comparing (28·22) with (15·3), we have

$$g_{11} = -e^\lambda, \quad g_{22} = -r^2, \quad g_{33} = -r^2 \sin^2\theta, \quad g_{44} = e^\nu . \; (28{\cdot}31)$$

and $\qquad g_{\sigma\tau} = 0$, when $\sigma \neq \tau$.

The determinant g reduces to its leading diagonal, so that

$$-g = e^{\lambda+\nu} r^4 \sin^2\theta, \quad . \quad . \quad . \quad . \quad . \quad (28{\cdot}32)$$

and

$$g^{\sigma\sigma} = 1/g_{\sigma\sigma} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (28{\cdot}33)$$

We can now calculate the three-index symbols (21·12)

$$\{\sigma\tau, \alpha\} = \tfrac{1}{2} g^{\alpha\beta} \left\{ \frac{\partial g_{\beta\sigma}}{\partial x_\tau} + \frac{\partial g_{\beta\tau}}{\partial x_\sigma} - \frac{\partial g_{\sigma\tau}}{\partial x_\beta} \right\}.$$

Since the g's vanish except when the two suffixes agree, the summation disappears and we have

$$\{\sigma\tau, \alpha\} = \frac{1}{2g_{\alpha\alpha}} \left\{ \frac{\partial g_{\alpha\sigma}}{\partial x_\tau} + \frac{\partial g_{\alpha\tau}}{\partial x_\sigma} - \frac{\partial g_{\sigma\tau}}{\partial x_\alpha} \right\} \quad \text{not summed.}$$

If σ, τ, ρ are unequal we get the following possible cases:—

$$\begin{aligned} \{\sigma\sigma, \sigma\} &= \frac{1}{2} \frac{\partial}{\partial x_\sigma} \log g_{\sigma\sigma} \\ \{\sigma\sigma, \tau\} &= -\frac{1}{2g_{\tau\tau}} \frac{\partial}{\partial x_\tau} g_{\sigma\sigma} \\ \{\sigma\tau, \tau\} &= \frac{1}{2} \frac{\partial}{\partial x_\sigma} \log g_{\tau\tau} \\ \{\sigma\tau, \rho\} &= 0. \end{aligned} \qquad (28{\cdot}4)$$

None of the above expressions are to be summed.

Whence by (28·31), denoting differentiation with respect to r by accents, we obtain

$$\left.\begin{aligned}
\{11, 1\} &= \tfrac{1}{2}\lambda' \\
\{12, 2\} &= 1/r \\
\{13, 3\} &= 1/r \\
\{14, 4\} &= \tfrac{1}{2}\nu' \\
\{22, 1\} &= -re^{-\lambda} \\
\{23, 3\} &= \cot\theta \\
\{33, 1\} &= -r\sin^2\theta e^{-\lambda} \\
\{33, 2\} &= -\sin\theta\cos\theta \\
\{44, 1\} &= \tfrac{1}{2}e^{\nu-\lambda}\nu'
\end{aligned}\right\} . \qquad (28\cdot5)$$

The remaining 31 Christoffel symbols are zero. It should be noted that $\{21, 2\}$ is the same as $\{12, 2\}$, etc.

It is now not difficult to obtain the equations of the field. To assist the reader in carrying through the substitutions, we shall write out in full the equations (28·1) omitting the terms (223 in number), which obviously vanish. The following come respectively from $G_{11}, G_{22}, G_{33}, G_{44}=0$:—

$$-\frac{\partial}{\partial r}\{11, 1\}+\{11, 1\}\{11, 1\}+\{12, 2\}\{12, 2\}+\{13, 3\}\{13, 3\}$$

$$+\{14, 4\}\{14, 4\}+\frac{\partial^2}{\partial r^2}\log\sqrt{-g}-\{11, 1\}\frac{\partial}{\partial r}\log\sqrt{-g}=0$$

$$-\frac{\partial}{\partial r}\{22, 1\}+2\{22, 1\}\{12, 2\}+\{23, 3\}\{23, 3\}+\frac{\partial^2}{\partial\theta^2}\log\sqrt{-g}$$

$$-\{22, 1\}\frac{\partial}{\partial r}\log\sqrt{-g}=0$$

$$-\frac{\partial}{\partial r}\{33, 1\}-\frac{\partial}{\partial\theta}\{33, 2\}+2\{33, 1\}\{13, 3\}+2\{33, 2\}\{23, 3\}$$

$$-\{33, 1\}\frac{\partial}{\partial r}\log\sqrt{-g}-\{33, 2\}\frac{\partial}{\partial\theta}\log\sqrt{-g}=0$$

$$-\frac{\partial}{\partial r}\{44, 1\}+2\{44, 1\}\{14, 4\}-\{44, 1\}\frac{\partial}{\partial r}\log\sqrt{-g}=0.$$

Of the remaining equations, $G_{12}=0$ gives

$$\{13, 3\}\{23, 3\}-\{12, 2\}\frac{\partial}{\partial\theta}\log\sqrt{-g}=0,$$

which disappears when the values of the symbols are substituted; and in the others there are no surviving terms.

Substituting from (28·5) and (28·32) the four equations give immediately

$$-\frac{1}{2}\lambda''+\frac{1}{4}\lambda'^2+\frac{1}{r^2}+\frac{1}{r^2}+\frac{1}{4}\nu'^2+\left(\frac{1}{2}\lambda''+\frac{1}{2}\nu''-\frac{2}{r^2}\right)$$
$$-\frac{1}{2}\lambda'\left(\frac{1}{2}\lambda'+\frac{1}{2}\nu'+\frac{2}{r}\right)=0,$$

$$e^{-\lambda}(1-r\lambda')-2e^{-\lambda}+\cot^2\theta-\operatorname{cosec}^2\theta+re^{-\lambda}\left(\frac{1}{2}\lambda'+\frac{1}{2}\nu'+\frac{2}{r}\right)=0,$$

$$\sin^2\theta e^{-\lambda}(1-r\lambda')+(\cos^2\theta-\sin^2\theta)-2\sin^2\theta e^{-\lambda}-2\cos^2\theta$$
$$+r\sin^2\theta e^{-\lambda}\left(\frac{1}{2}\lambda'+\frac{1}{2}\nu'+\frac{2}{r}\right)+\cos^2\theta=0$$

$$-\frac{1}{2}e^{\nu-\lambda}(\nu''+\nu'^2-\nu'\lambda')+\frac{1}{2}e^{\nu-\lambda}\nu'^2-\frac{1}{2}e^{\nu-\lambda}\nu'\left(\frac{1}{2}\lambda'+\frac{1}{2}\nu'+\frac{2}{r}\right)=0.$$

These reduce to

$$\left.\begin{aligned}
G_{11}&=\tfrac{1}{2}\nu''-\tfrac{1}{4}\lambda'\nu'+\tfrac{1}{4}\nu'^2-\lambda'/r=0,\\
G_{22}&=e^{-\lambda}(1+\tfrac{1}{2}r(\nu'-\lambda'))-1=0,\\
G_{33}&=\sin^2\theta\,.\,e^{-\lambda}(1+\tfrac{1}{2}r(\nu'-\lambda'))-\sin^2\theta=0,\\
G_{44}&=e^{\nu-\lambda}(-\tfrac{1}{2}\nu''+\tfrac{1}{4}\lambda'\nu'-\tfrac{1}{4}\nu'^2-\nu'/r)=0,
\end{aligned}\right\} \quad . \quad (28{\cdot}6)$$

From the first and last equations $\lambda'=-\nu'$, and since both λ and ν must tend to zero at infinity $\lambda=-\nu$. The second and third equations (which are identical) then give

$$e^{\nu}(1+r\nu')=1.$$

Set $e^{\nu}=\gamma$, then $\gamma+r\gamma'=1.$

Whence
$$\gamma=1-\frac{2m}{r}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (28{\cdot}7)$$

where $2m$ is a constant of integration. m will later be identified with the mass of the particle in gravitational units. This solution satisfies the first and fourth equations, and, therefore, substituting in (28·22), we have as a possible expression for the line-element

$$ds^2=-\gamma^{-1}dr^2-r^2d\theta^2-r^2\sin^2\theta d\varphi^2+\gamma dt^2, \quad . \quad (28{\cdot}8)$$

with $\gamma=1-2m/r$.

It will be seen that the measured space around a particle is not Euclidean. Any actual measurement with our clock-scale gives the invariant quantity ds. If we lay our measuring-rod transversely, $ds=rd\theta$, so that our transverse measures are

correct in this system of co-ordinates; but if we lay it radially, $ds = \gamma^{-\frac{1}{2}}dr$, and the measures need to be multiplied by $\gamma^{\frac{1}{2}}$ to give dr. Thus, referring our results to Euclidean space, we may say that a standard measuring rod contracts when turned from the transverse to the radial direction.

We could, of course, decide to treat the radial measures as correct, and apply corrections to the transverse measures. This amounts to substituting dr' for $\gamma^{-\frac{1}{2}}dr$ in (28·8), and using r' as the radial co-ordinate. It is impossible to say which form of (28·8) corresponds to our ordinary polar co-ordinates, since we have never hitherto had to pay attention to the ambiguity.

The possibility of using any function of r, instead of r, for the distance is connected with the fact that Einstein's equations amount to only 6 independent relations between the 10 g's. Consequently, quite apart from boundary conditions, there is a large amount of arbitrariness in choice of g's, *i.e.*, of co-ordinates. The reader may meet elsewhere with different expressions for the line-element due to a particle. The one adopted here was first given by Schwarzschild.

For some purposes the following analogy is helpful. Instead of considering continuous space-time, consider that fundamentally we are dealing with an aggregate of points. With Galilean co-ordinates x, y, z, $t\sqrt{-1}$ the points are uniformly packed. Any measure that we make is really a counting of points, and a particle always moves so as to pass through the fewest possible points between any two positions on its path. Any mathematical transformation of these co-ordinates disturbs, without disordering, the distribution of the points in space; but it is meaningless so long as we consider only the points and not the arbitrary continuous space we place them in. In a gravitational field the points are disordered according to some definite law. We can evidently re-arrange them so that the number of points in the circumference of a circle is less than π times the number in the diameter (a circle being a geodesic on a hypersphere, which is a locus such that the minimum number of points between any point on it and a fixed point called the centre is constant).

This representation, however, gives only imaginary time and therefore imaginary motions. When extended to real motions it becomes too complex to be of much help.

CHAPTER V.

THE CRUCIAL PHENOMENA.

29. *The Equations of Motion of a Particle in the Gravitational Field.*

Denote the contravariant vector $\frac{\partial x_\sigma}{\partial s}$ by A^σ. Then by (22·5) its covariant derivative is

$$A^\sigma_\alpha = \frac{\partial}{\partial x_\alpha}\left(\frac{\partial x_\sigma}{\partial s}\right) + \{\alpha\beta, \sigma\} \frac{\partial x_\beta}{\partial s}.$$

Multiply this by $\partial x_\alpha / \partial s$, we have

$$A^\alpha A^\sigma_\alpha = \frac{\partial^2 x_\sigma}{\partial s^2} + \{\alpha\beta, \sigma\} \frac{\partial x_\alpha}{\partial s} \frac{\partial x_\beta}{\partial s},$$

showing that the right-hand side is a contravariant vector.

Consider the equations

$$\frac{\partial^2 x_\sigma}{\partial s^2} + \{\alpha\beta, \sigma\} \frac{\partial x_\alpha}{\partial s} \frac{\partial x_\beta}{\partial s} = 0, \quad (\sigma = 1, 2, 3, 4), \quad . \quad . \quad (29)$$

since the left-side is a vector, the equations will be satisfied (or not) independently of the choice of co-ordinates. In Galilean co-ordinates, the second term vanishes, and the equations reduce to $\partial^2 x_\sigma / \partial s^2 = 0$, which are the equations of a straight line. Equation (29) is thus the general equation of the locus which in Galilean co-ordinates becomes a straight line.

The path of a particle in Galilean co-ordinates (*i.e.*, under no forces) is a straight line. The equations (29) are accordingly the equations of motion of a particle referred to any axes, provided there is no permanent gravitational field. Further, since they contain only first derivatives of the g's, in accordance with § 27, these equations of motion will hold also when there is a permanent gravitational field.

The equations must evidently correspond to the condition,

$$\int ds \text{ is stationary,}$$

and could have been deduced from it by the calculus of variations. The path of a particle is a geodesic in all cases. It should be noticed that $\int ds$ is not generally a *minimum*.

30. Using the values (28·5) of Christoffel's symbols, the equation of motion (29) for $\sigma=2$ becomes

$$\frac{d^2\theta}{ds^2}-\cos\theta\sin\theta\left(\frac{d\varphi}{ds}\right)^2+\frac{2}{r}\frac{dr}{ds}\frac{d\theta}{ds}=0. \quad . \quad (30\cdot12)$$

Choose co-ordinates so that the particle moves initially in the plane $\theta=\pi/2$; then $d\theta/ds=0$ initially, and $\cos\theta=0$, so that $d^2\theta/ds^2=0$. The particle therefore continues to move in this plane. The equations for $\sigma=1, 3, 4$ are then

$$\frac{d^2r}{ds^2}+\tfrac{1}{2}\lambda'\left(\frac{dr}{ds}\right)^2-re^{-\lambda}\left(\frac{d\varphi}{ds}\right)^2+\tfrac{1}{2}e^{\nu-\lambda}\nu'\left(\frac{dt}{ds}\right)^2=0 \quad (30\cdot11)$$

$$\frac{d^2\varphi}{ds^2}+\frac{2}{r}\frac{dr}{ds}\frac{d\varphi}{ds}=0 \quad (30\cdot13)$$

$$\frac{d^2t}{ds^2}+\nu'\frac{dr}{ds}\frac{dt}{ds}=0 \quad (30\cdot14)$$

Integrating (30·13) and (30·14), we have

$$r^2\frac{d\varphi}{ds}=h, \quad . \quad . \quad . \quad (30\cdot21)$$

$$\frac{dt}{ds}=ce^{-\nu}=\frac{c}{\gamma}, \quad (30\cdot22)$$

where h and c are constants of integration.

Instead of troubling to integrate (30·11), we can use (28·8), which plays the part of an integral of energy, viz.,

$$\gamma^{-1}\left(\frac{dr}{ds}\right)^2+r^2\left(\frac{d\varphi}{ds}\right)^2-\gamma\left(\frac{dt}{ds}\right)^2=-1 \quad . \quad . \quad (30\cdot23)$$

From these three integrals,

$$\left(\frac{dr}{ds}\right)^2+r^2\left(\frac{d\varphi}{ds}\right)^2+(\gamma-1)\frac{h^2}{r^2}-c^2=-\gamma,$$

or substituting for γ its value (28·7)

$$\left.\begin{aligned}\left(\frac{dr}{ds}\right)^2+r^2\left(\frac{d\varphi}{ds}\right)^2&=(c^2-1)+\frac{2m}{r}+2m\frac{h^2}{r^3}\\ \text{with}\qquad r^2\frac{d\varphi}{ds}&=h\end{aligned}\right\}\quad(30\cdot3)$$

Compare these with the ordinary Newtonian equations for elliptic motion,

$$\left.\begin{aligned}\left(\frac{dr}{dt}\right)^2+r^2\left(\frac{d\varphi}{dt}\right)^2&=-\frac{m}{a}+\frac{2m}{r},\\ r^2\frac{d\varphi}{dt}&=h.\end{aligned}\right\} \quad (30{\cdot}4)$$

To make them correspond we must take $c^2=1-m/a$, where a is the major semiaxis of the orbit. The term $2mh^2/r^3$ represents a small additional effect not predicted by the Newtonian theory. Further, the quantity m, introduced as a constant of integration, is now identified as the mass of the attracting particle measured in gravitational units. With regard to the use of ds instead of dt in (30·3), it must be remembered that ds is the "proper time" for the moving particle, so it is permissible to take ds as corresponding to the time in making a comparison with Newtonian dynamics.

Mass, time and distance are all ambiguously defined in Newtonian dynamics, and in defining them for the present theory we have some freedom of choice, provided that our definition agrees with the Newtonian definition in the limiting case of a vanishing field of force.

31. *The Perihelion of Mercury.*

The ratio m/a or m/r is very small in all practical applications. If we take 1 kilometre as the unit of length and time $\left(=\frac{1}{300{,}000}\text{ sec.}\right)$, then for the earth's orbit $a=1{\cdot}49\,.\,10^8$, and the angular velocity $\omega=6{\cdot}64\,.\,10^{-13}$. Hence the mass of the sun,

$$m=\omega^2a^3=1{\cdot}47 \text{ kilometres} \quad . \quad . \quad . \quad (31{\cdot}1)$$

Thus for applications in the solar system m/r is of order 10^{-8} and it is easily seen that h^2/r^2 is of the same order. Also the difference between dt and ds is of order $10^{-8}ds$.

From (30·3) we have

$$\left(\frac{h}{r^2}\frac{dr}{d\varphi}\right)^2+\frac{h^2}{r^2}=(c^2-1)+\frac{2m}{r}+\frac{2mh^2}{r^3},$$

or writing $u=1/r$.

$$\left(\frac{du}{d\varphi}\right)^2+u^2=\frac{c^2-1}{h^2}+\frac{2mu}{h^2}+2mu^3.$$

Differentiating with respect to φ,

$$\frac{d^2u}{d\varphi^2}+u=\frac{m}{h^2}+3mu^2 \quad . \quad . \quad . \quad . \quad . \quad (31\cdot 2)$$

Since h^2u^2 is of order 10^{-8} we obtain an approximate solution by neglecting $3mu^2$. This is

$$u=\frac{m}{h^2}(1+e\cos(\varphi-\varpi)), \quad . \quad . \quad . \quad (31\cdot 3)$$

as in Newtonian dynamics.

For a second approximation, we substitute this value of u in the small term $3mu^2$, and (31·2) becomes

$$\frac{d^2u}{d\varphi^2}+u=\frac{m}{h^2}+\frac{3m^3}{h^4}+\frac{6m^3}{h^4}e\cos(\varphi-\varpi)+\frac{3m^3e^2}{2h^4}(1+\cos 2(\varphi-\varpi)). \quad (31\cdot 4)$$

Of the small additional terms the only one which can give appreciable effects is the term in $\cos(\varphi-\varpi)$, which is of the proper period to produce a continually increasing effect by resonance. It is well known that the particular integral of

$$\frac{d^2u}{d\varphi^2}+u=A\cos\varphi$$

is

$$u=\tfrac{1}{2}A\varphi\sin\varphi.$$

Hence this term gives a part of u,

$$u_1=\frac{3m^3e}{h^4}\varphi\sin(\varphi-\varpi).$$

Adding this to (31·3) we have

$$u=\frac{m}{h^2}\left(1+e\cos(\varphi-\varpi)+\frac{3m^2}{h^2}\varphi\, e\sin(\varphi-\varpi)\right)$$

$$=\frac{m}{h^2}(1+e\cos(\varphi-\varpi-\delta\varpi)),$$

where $\delta\varpi=\frac{3m^2}{h^2}\varphi$, and $(\delta\varpi)^2$ is neglected.

Thus whilst the planet moves through one revolution, the perihelion advances a fraction of a revolution equal to

$$\frac{\delta\varpi}{\varphi}=\frac{3m^2}{h^2}=\frac{3m}{a(1-e^2)}=\frac{12\pi^2a^2}{c^2T^2(1-e^2)},$$

where T is the period of the planet, and the velocity of light c has been re-instated.

For the four inner planets the numerical values of this predicted motion of the perihelion are (per century) :—

	$\delta\varpi$.	$e\delta\varpi$.
Mercury	$+42''\cdot9$	$+8''\cdot82$
Venus	8·6	0·05
Earth	3·8	0·07
Mars	1·35	0·13

The value of $e\delta\varpi$ is given because this corresponds to the perturbation which can be measured. Clearly when e is vanishingly small it is not possible to detect observationally any change in the position of perihelion. The orbits of Venus and the Earth are nearly circular so that the predicted effect is too small to detect.

The following table gives the outstanding discrepancies between the present theory and observation for $e\delta\varpi$ and δe (per century) with their probable errors. The secular changes δe are analogous to $e\delta\varpi$; and the two perturbations may be regarded as the two rectangular components of a vector. In the last column we give the outstanding discrepancies of $e\delta\varpi$ on the Newtonian theory ; those of δe are, of course, unaltered.

	Einstein's Theory.		Newtonian.
	$e\delta\varpi$	δe	$e\delta\varpi$
Mercury	$-0''\cdot58 \pm 0''\cdot29$	$-0''\cdot88 \pm 0\cdot''33$	$+8''\cdot24$
Venus	$-0\cdot11 \pm 0\cdot17$	$+0\cdot21 \pm 0\cdot21$	$-0\cdot06$
Earth	$0\cdot00 \pm 0\cdot09$	$+0\cdot02 \pm 0\cdot07$	$+0\cdot07$
Mars	$+0\cdot51 \pm 0\cdot23$	$+0\cdot29 \pm 0\cdot18$	$+0\cdot64$

It will be seen that the famous large discordance of the perihelion of Mercury is removed by Einstein's theory. No other charge of importance is made except a slight improvement, for the perihelion of Mars. Of the eight residuals, four exceed the probable error, and none exceed three times the probable error, so that the agreement is very satisfactory.

It may be noticed that according to (31·4) the orbit is not exactly an ellipse, even apart from this progression of the apse. But this (unlike the motion of perihelion) has no observational significance, and merely arises from our particular choice of measurement of r. In any case the curve in non-Euclidean space, which is to be described as an ellipse, must be a matter of convention.

It will be found (putting $dr/ds=0$ in (30·11)) that for a circular orbit Kepler's third law is exactly fulfilled. This again is not an observable fact. To compare it with observation we should have to consider the nature of the astronomical observations from which the direct value of the axis of the orbit is measured.

32. *Deflection of a Ray of Light.*

In the absence of a gravitational field the velocity of light is unity, so that

$$\left(\frac{dx}{dt}\right)^2+\left(\frac{dy}{dt}\right)^2+\left(\frac{dz}{dt}\right)^2=1.$$

Accordingly $ds^2=-dx^2-dy^2-dz^2+dt^2=0$. . . . (32·1)

Hence for the motion of light $ds=0$, and by the principle of equivalence this invariant equation must hold also in the gravitational field.

It may be of interest to notice that for an observer travelling with the light, $dx=dy=dz=0$, so that $dt=ds=0$. Hence, if man wishes to achieve immortality and eternal youth, all he has to do is to cruise about space with the velocity of light. He will return to the earth after what seems to him an instant to find many centuries passed away.

Setting $ds=0$ in (28·8) we have (for motion in a plane)

$$\gamma^{-1}\left(\frac{dr}{dt}\right)^2+\left(r\frac{d\phi}{dt}\right)^2=\gamma \quad . \quad . \quad . \quad . \quad . \quad (32·2)$$

Hence if v is the velocity of light in a direction making an angle V with the radius vector,

$$v^2(\gamma^{-1}\cos^2 V+\sin^2 V)=\gamma,$$

whence $$v=\gamma(1-(1-\gamma)\sin^2 V)^{-\frac{1}{2}} \quad . \quad . \quad . \quad . \quad (32·3)$$

The velocity thus depends on the direction; but it must be remembered that this co-ordinate velocity is not the velocity found directly from measures at the point considered. When we determine the velocity by measures made in a small region, and use natural measure (*i.e.*, $g_{\mu\nu}$ having the values (16·3) at that point), the measured velocity is necessarily unity.

Since it is inconvenient to have the velocity of light varying with direction, we shall slightly alter our co-ordinates. Set

$$r=r_1+m \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (32·4)$$

Then, neglecting squares of m/r_1,

$$r^2=r_1^2(1+2m/r)=\gamma^{-1}r_1^2.$$

Substituting in (32·2)

$$\left(\frac{dr_1}{dt}\right)^2+\left(r_1\frac{d\phi}{dt}\right)^2=\gamma^2,$$

so that in these co-ordinates,

$$v=\gamma=1-2m/r\simeq 1-2m/r_1 \quad . \quad . \quad . \quad . \quad (32·5)$$

for all directions. We can now drop the suffix of r_1.

By Huygens' principle the direction of the ray is determined by the condition that the time between two points is stationary for small variations of the path. The course of the ray will therefore depend only on the variation of velocity, and will be the same as in a Euclidean space filled with material of suitable refractive index. The necessary refractive index μ is given by

$$\mu=\frac{1}{v}=1+\frac{2m}{r}. \quad . \quad . \quad . \quad . \quad . \quad (32·6)$$

We thus see that the gravitational field round a particle will act like a converging lens.

The path of a ray through a medium stratified in concentric spheres is given by

$$\mu p=\text{const.} \quad . \quad . \quad . \quad . \quad . \quad . \quad (32·71)$$

where p is the perpendicular from the centre on the tangent.

By (32·6) we have to this order of approximation,

$$\mu^2=1+\frac{4m}{r}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (32·72)$$

But (32·71) and (32·72) are the integrals of angular momentum and energy for the Newtonian motion of a particle with velocity μ under the attraction of a mass $2m$, the orbit being a hyperbola of semi-axis $2m$. This hyperbola, therefore, gives the path of the light. If the distance from the focus to the apse is R, we have

$$a=2m,$$
$$a(e-1)=R,$$

so that

$$e=1+\frac{R}{2m}\simeq\frac{R}{2m},$$

and the very small angle between the asymptotes

$$=\frac{2}{\sqrt{(e^2-1)}}\simeq\frac{2}{e}\simeq\frac{4m}{R}.$$

Thus a ray of light travelling from $-\infty$ to $+\infty$, and passing at a distance R from a particle of mass m experiences a total deflection.

$$\alpha=\frac{4m}{R}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (32·8)$$

For a star seen close to the limb of the sun, by (31·1) $m=1\cdot47$ km., and R=sun's radius=697,000 km. Hence

$$\alpha=1''\cdot74.$$

It is curious to notice the occurrence of the factor 2 (mass=$2m$) in the dynamical analogy. The deflection is twice what we should obtain on the Newtonian theory for a particle moving through the gravitational field with the velocity of light. The path of a light ray is not a geodesic (or rather the notion of a geodesic fails for motion with the speed of light); it is determined by stationary values of $\int dt$ instead of $\int ds$.

It may also be noted that the velocity of light decreases as the light falls to the attracting body.

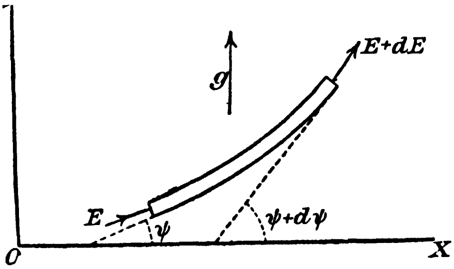

FIG. 4.

33. It is hoped to test this prediction by observations of stars near the limb of the sun during a total eclipse. If the answer should be in the affirmative, the question will arise whether this must be considered to confirm Einstein's law of gravitation, or whether the deflection is sufficiently accounted for by the simple hypothesis that the mass of the electromagnetic energy of light is subject to gravitation. The unexpected factor 2 suggests that the deflection on Einstein's theory will be double that which would result from the ordinary electromagnetic theory. It is worth while to examine this more closely.

Consider a tube of light of unit cross-section and length ds (Fig. 4). Let the inclination of the ray to the axis of x be ψ. Let g be the acceleration of the gravitational field directed along Oy. Let E be the energy per unit volume; and c be the velocity of light, which on the electromagnetic theory is absolutely constant.

Then the mass of electromagnetic energy E, according to electromagnetic theory (or by (7·85)), is E/c^2, so that if this is subject to gravity the momentum generated in the tube in unit time will be

$$\frac{E}{c^2}ds \, . \, g \quad \text{along } Oy.$$

If the light is stopped by an absorbing screen placed perpendicular to the ray the radiation-pressure is numerically equal to E, showing that momentum E in the direction of the ray passes across a section of the tube in unit time. Thus, resolving in the x and y directions, the conservation of momentum gives

$$\left.\begin{aligned} &\frac{d}{ds}(E\cos\psi) \, . \, ds=0, \\ &\frac{d}{ds}(E\sin\psi) \, . \, ds=\frac{gE}{c^2}ds, \end{aligned}\right\} \qquad (33{\cdot}1)$$

Whence

$$\left.\begin{aligned} &\frac{dE}{ds}\cos\psi - E\sin\psi\frac{d\psi}{ds}=0, \\ &\frac{dE}{ds}\sin\psi + E\cos\psi\frac{d\psi}{ds}=\frac{gE}{c^2}. \end{aligned}\right\} \qquad (33{\cdot}2)$$

Eliminating dE/ds,

$$\frac{d\psi}{ds}=\frac{g}{c^2}\cos\psi. \qquad (33{\cdot}3)$$

The radius of curvature $ds/d\psi$ is thus $c^2/g\cos\psi$, which is exactly the same as for a material particle moving with velocity c in ordinary dynamics. This, as shown in the last paragraph, is only half the deflection indicated by Einstein's theory; and the experimental amount of the deflection should thus provide a crucial test.

34. *Displacement of Spectral Lines.*

Consider an atom vibrating at any point of the gravitational field. It is a natural clock which ought to give an invariant measure of an interval δs; that is to say, the interval δs corresponding to one vibration of the atom is always the same. Let the atom be momentarily at rest in our system of coordinates (though subject to the acceleration of the field); then $dx=dy=dz=0$, and by (15·3)

$$ds^2=g_{44}dt^2.$$

If then dt and dt' are the periods of two simi ar atoms v.brating at different parts of the field where the potentials are g_{44} and g'_{44}, respectively,

$$\sqrt{g_{44}}.dt=\sqrt{g'_{44}}.dt'. \quad . \quad . \quad . \quad . \quad . \qquad (34\cdot 1)$$

If t refers to an atom vibrating in the photosphere of the sun,

$$g_{44}=1-\frac{2m}{R},$$

and if t' refers to an atom in a terrestrial laboratory, where g'_{44} is practically unity,

$$\frac{dt}{dt'} \simeq 1+\frac{m}{R}=1\cdot 00000212 \quad . \quad . \quad . \qquad (34\cdot 2)$$

The solar atom thus vibrates more slowly, and its spectral lines will be displaced towards the red. The amount is equivalent to the Doppler displacement due to a velocity of 0·00000212, or in ordinary units 0·634 km. per sec. In the part of the spectrum usually investigated the displacement is about 0·008 tenth-metres.

The effect is of particular importance, because it has been claimed that the existence of this displacement is disproved by observations of the solar spectrum.* The difficulties of the test are so great that we may perhaps suspend judgment; but it would be idle to deny the seriousness of this apparent break-down of Einstein's theory. We shall therefore consider the phenomenon from a more elementary point of view.

The phenomenon does not depend on the greater intensity of the field on the sun, but on the potential; and it can evidently occur in a uniform gravitational field. Consider an observer O in a uniform field of intensity g and two similar atoms A_1 and A_2, A_1 being close to the observer and A_2 at a distance a measured parallel to the field. The observer and his atoms will, of course, be falling with the acceleration g. Consider them all enclosed in a room which is also falling; then by the principle of equivalence O cannot detect any effect of the field, and he will therefore observe the same period of vibration T for both atoms. Now refer the phenomena to unaccelerated axes which coincide with the accelerated axes at the instant $t=0$. The vibration emitted by A_2 at the time $t=0$ will reach O at the time $t=a$ (the velocity of

* C. E. St. John, "Astrophysical Journal," Vol. 46, p. 249.

light being unity), by which time O will have acquired a velocity ga relative to the unaccelerated axes. He will, therefore, correct his observation of the period of A_2 for the Doppler effect of this velocity and deduce a true period $T/(1-ga)$. The period of A_1 will require no correction, and wil still be given as T. Since ga is the difference of potential between A_1 and A_2 this agrees with (34·2).

As an example of a varying field, consider an observer O at the origin of co-ordinates and an atom A at a distance r in a field of centrifugal force of potential $\Omega=\frac{1}{2}\omega^2r^2$, the atom being at rest at the time of emission of the light, but subject to the acceleration of the field. Another way of stating the problem is that there is no field of force, and the atom is moving with velocity ωr at right angles to the radius vector at the time of emission of the light. But in that case the period of vibration is by § 4 increased in the ratio

$$\begin{aligned}\beta&=(1-\omega^2r^2)^{-\frac{1}{2}}=(1-2\Omega)^{-\frac{1}{2}}\\&=1/\sqrt{g_{44}}\quad\text{by (16·2),}\end{aligned}$$

as compared with the stationary atom. This again agrees with (34·1).

These verifications seem to leave little chance of evading the conclusion that a displacement of the Fraunhofer lines is a necessary and fundamental condition for the acceptance of Einstein's theory; and that if it is really non-existent, under conditions which strictly accord with those here postulated, we should have to reject the whole theory constructed on the principle of equivalence. Possibly a compromise might be effected by supposing that gravitation is an attribute only of matter in bulk and not of individual atoms; but this would involve a fundamental restatement of the whole theory.

If the displacement of the solar lines were confirmed, it would be the first *experimental* evidence that relativity holds for quantum phenomena.

CHAPTER VI.

THE GRAVITATION OF A CONTINUOUS DISTRIBUTION OF MATTER.

35. In the problems occurring in Nature our data give, not the distribution of the individual atoms, but the large-scale average distribution of density. This transition from discrete particles to the equivalent continuous medium occurs in the Newtonian theory of attractions, and involves the replacement of Laplace's equation $\nabla^2\varphi=0$ by Poisson's equation $\nabla^2\varphi=-4\pi\rho$. We shall now find the corresponding modification of Einstein's equations $G_{\sigma\tau}=0$.

The equations $G_{\sigma\tau}=0$ are not linear in the g's, and consequently the fields of two or more particles are not strictly additive. But the deviations produced in the g's by any natural gravitational field are extremely small, so we shall neglect the product terms and treat the fields as superposable. It will be shown below that ultimately this approximation does not produce any inaccuracy in the application we have in view.

As in (32·4) we shall write $r=r_1+m$ in (28·8) and neglect $(m/r)^2$. Then the line element in the field surrounding the particle is

$$ds^2=-\gamma^{-1}(dr_1{}^2+r_1{}^2d\theta^2+r_1{}^2\sin^2\theta d\varphi^2)+\gamma dt^2. \quad . \quad (35\cdot1)$$

We consider r_1 to be the actual radius vector, since the mode of measurement is arbitrary to this extent. Converting into rectangular co-ordinates,

$$\begin{aligned} ds^2&=-\gamma^{-1}(dx^2+dy^2+dz^2)+\gamma dt^2,\\ &=-\left(1+\frac{2m}{r}\right)(dx^2+dy^2+dz^2)+\left(1-\frac{2m}{r}\right)dt^2 \quad . \quad (35\cdot2)\end{aligned}$$

The origin is now arbitrary, and r denotes the distance of the attracting particle from the element ds. The effects of a number of particles being additive to our order of approxi-

mation, we shall have for any number of particles at rest relative to the axes,

$$ds^2 = -(1+2\Omega)(dx^2+dy^2+dz^2)+(1-2\Omega)dt^2 \quad . \quad (35\cdot3)$$

where $\Omega=\Sigma(m/r)$=the Newtonian potential.

Consider a point O in the medium where the density is ρ, and with O as centre describe an infinitely small sphere. If we neglect the material inside the sphere, the equations of the gravitational field in free space will be satisfied at O, *i.e.*, $G_{\sigma\tau}=0$. Hence in calculating the values of $G_{\sigma\tau}$ at O we need only take account of the material inside the sphere. Accordingly in (35·3) Ω refers to the potential inside an infinitely small sphere of uniform density ρ.

Since $\partial\Omega/\partial x$, &c., vanish at O, we have only to take account of terms in (28·1) containing second derivatives of the g's; and the calculation of $G_{\sigma\tau}$ at O is quite simple. We have

$$\begin{aligned} G_{11} &= -\frac{\partial}{\partial x_1}\{11, 1\} - \frac{\partial}{\partial x_2}\{11, 2\} - \frac{\partial}{\partial x_3}\{11, 3\} + \frac{\partial^2}{\partial x_1^2}\log\sqrt{-g} \\ &= -\tfrac{1}{2}g^{11}\frac{\partial^2 g_{11}}{\partial x_1^2} + \tfrac{1}{2}g^{22}\frac{\partial^2 g_{11}}{\partial x_2^2} + \tfrac{1}{2}g^{33}\frac{\partial^2 g_{11}}{\partial x_3^2} + \frac{\partial^2}{\partial x_1^2}\log\sqrt{-g} \quad . \quad (35\cdot4) \end{aligned}$$

omitting 33 terms which vanish or cancel.

At O, $\qquad g_{11}=g^{11}=-1, \quad g_{44}=g^{44}=1, \quad . \quad . \quad (35\cdot5)$

and by (35·3)

$$\frac{\partial^2 g_{11}}{\partial x_1^2} = \frac{\partial^2 g_{44}}{\partial x_1^2} = -\frac{\partial^2}{\partial x_1^2}\log\sqrt{-g} = -2\frac{\partial^2\Omega}{\partial x_1^2}.$$

Hence substituting in (35·4)

$$\begin{aligned} G_{11} &= \frac{\partial^2\Omega}{\partial x_1^2} + \frac{\partial^2\Omega}{\partial x_2^2} + \frac{\partial^2\Omega}{\partial x_3^2}, \\ &= -4\pi\rho, \quad \text{by Poisson's equation.} \end{aligned}$$

Working out the other components similarly (with slight variations in the case of G_{44}) we find

$$G_{11}=G_{22}=G_{33}=G_{44}=-4\pi\rho. \quad . \quad . \quad (35\cdot6)$$

The scalar

$$\begin{aligned} G &= g^{\sigma\tau}G_{\sigma\tau}, \\ &= -G_{11}-G_{22}-G_{33}+G_{44}, \\ &= 8\pi\rho. \quad . \quad . \quad . \quad . \quad . \quad (35\cdot7) \end{aligned}$$

Now form the covariant tensor

$$-8\pi T_{\sigma\tau} = G_{\sigma\tau} - \tfrac{1}{2}g_{\sigma\tau}G. \quad . \quad . \quad . \quad . \quad (35\cdot8)$$

We have by (35·6) and (35·7)

$$T_{44}=\rho,$$

and all other components vanish.

Having thus found the value of $T_{\sigma\tau}$ in this special system of co-ordinates we could find its general value by (19·31). It is, however, simpler to proceed as fo lows. If x_μ is a co-ordinate of a point in the material, consider the quantity,

$$\rho\frac{dx_\mu}{ds}\frac{dx_\nu}{ds}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (35\cdot91)$$

Since with respect to our special axes the material is at rest,

$$\frac{dx_\mu}{ds}=0 \quad (\mu=1, 2, 3), \quad \text{and} \quad \frac{dx_\mu}{ds}=1 \quad (\mu=4).$$

Hence all the components of (35·91) vanish except for $\mu=\nu=4$, for which the component is ρ—just like $T_{\sigma\tau}$. This, however, is a contravariant tensor * and (35·8) requires a covariant tensor.

We therefore form the associated covariant tensor (§ 20*b*)

$$T_{\sigma\tau}=\rho\cdot g_{\mu\sigma}g_{\nu\tau}\frac{dx_\mu}{ds}\frac{dx_\nu}{ds}, \quad . \quad . \quad . \quad (35\cdot92)$$

which agrees with (35·91) in our special co-ordinates.

The equations (35·8) and (35·92) are in covariant form, and are true in one system, hence they are true in all possible systems of co-ordinates. They are the general equations of the gravitational field in a continuous medium.

An alternative form of (35·8) is readily obtained, viz.,

$$G_{\sigma\tau}=-8\pi(T_{\sigma\tau}-\tfrac{1}{2}g_{\sigma\tau}T), \quad . \quad . \quad . \quad (35\cdot93$$

where T is the associated scalar $g^{\sigma\tau}T_{\sigma\tau}$. (This follows since on inner multiplication of (35·8) by $g^{\sigma\tau}$ we obtain $G=8\pi T$.)

36. We thus find that in a continuous medium, $G_{\sigma\tau}$, instead of vanishing, is equal to a tensor expressing the content and state of motion of the medium at the point considered. On the equations here found we have two observations to make.

(1) A little consideration will show that notwithstanding the approximations made at various stages of the proof, the results are quite rigorous. It is clear that so far as the calculations for the infinitely small sphere surrounding O are concerned,

* ρ is to be treated as an invariant. Whatever the axes chosen, ρ is to be the density in natural measure as estimated by an observer moving with the matter.

we are justified in neglecting the product terms, since in the limit they will vanish compared with the linear terms. Another way of seeing this is to consider that $G_{\sigma\tau}$ involves only derivatives up to the second at the origin; and therefore we need only expand the g's in powers of r as far as r^2; but in our units ρ is of dimensions r^{-2}, and since the g's in rectangular co-ordinates are of zero dimensions, any terms involving ρ^2 would be of the form $\rho^2 r^4$, and therefore need not be retained. The effect of the gravitation of the matter outside the sphere is eliminated completely by our choice of co-ordinates. We chose them so that at O the g's have the values (16·3), *i.e.*, we use "natural measure." Since our axes move with the matter at O, the first derivatives of the g's (expressing the force) will not vanish unless the matter at O is moving with the acceleration of the field, which is not the case if there is any internal stress. These first derivatives are omitted from our equations after (35·3), because as already explained the external matter alone contributes nothing to $G_{\sigma\tau}$; further, the cross-terms are zero, because the first derivatives of the g's arising from the matter inside the sphere vanish. The result is thus rigorous, provided that in measuring the invariant density ρ we use natural measure, *i.e.*, the mass and unit volume must be taken according to the direct measures made by an observer at O moving with the material there.

The argument may be summarised thus: $G_{\sigma\tau}$ consists of terms of types

$$I_2+E_2+I_1^2+I_1E_1+E_1^2+\text{terms in } I_0+\text{terms in } E_0,$$

where I and E refer to the matter internal and external to the small sphere, and the suffixes refer to the order of the derivatives. Terms in I_1 vanish by the symmetry of the sphere; terms in I_0 vanish as the sphere is made infinitely small; terms in E_0 vanish because we use natural measure; the terms $E_2+E_1^2$ vanish by Einstein's equations for free space. All that is left is I_2, and as the sphere is made infinitely small our determination of its value becomes rigorous.

(2) In replacing a molecular medium by a continuous medium, it is not sufficient to average the distribution of mass and mass-motion only; we must also represent somehow the internal motions. This is done by adding another property to the continuous medium—the pressure, or stress-system. The tensor $T_{\sigma\tau}$ will contain terms corresponding to the pressure; these are negligible in practical calculations of the gravitational

field because the pressure is of order ρ times the Newtonian potential, *i.e.*, of order ρ^2. The terms are, however, important in the general equations of momentum and energy, and we shall consider them more fully in the next paragraph.

37. In the dynamics of a continuous medium the most fundamental part is taken by the associated mixed tensor,

$$T_\mu^\nu = g^{\nu\alpha} T_{\mu\alpha} = g_{\mu\sigma} \Sigma \rho_0 \frac{dx_\sigma}{ds} \frac{dx_\nu}{ds}, \quad . \quad . \quad . \quad (37{\cdot}1)$$

where we have inserted the Σ in order to take account of the variety of internal motions, and have written ρ_0 for ρ in order to call attention to the fact that it represents the density in natural measure and not the density referred to the arbitrary axes chosen.

T_μ^ν may be called the energy-tensor, though it is actually an *omnium gatherum* of energy, mass, stress and momentum.

First consider the meaning of this tensor in the absence of a gravitational field, and accordingly choose Galilean axes. If u, v, w are the component velocities of the particles,

$$\frac{dx}{dt} = u, \quad \frac{dy}{dt} = v, \quad \frac{dz}{dt} = w, \quad \left(\frac{ds}{dt}\right)^2 = 1 - u^2 - v^2 - w^2 = \beta^{-2}. \quad (37{\cdot}2)$$

But by (7·92) the density referred to the axes chosen is

$$\rho = \beta^2 \rho_0 = \rho_0 \Big/ \left(\frac{ds}{dt}\right)^2.$$

Hence
$$T_\mu^\nu = g_{\mu\sigma} \Sigma \rho \frac{dx_\sigma}{dt} \frac{dx_\nu}{dt}. \quad . \quad . \quad . \quad . \quad . \quad (37{\cdot}3)$$

Putting in the Galilean values of $g_{\mu\sigma}$, we have

$$\begin{array}{c} \xrightarrow{\quad\mu\quad} \\ \downarrow \nu \end{array} \quad T_\mu^\nu = \begin{array}{llll} -\Sigma\rho u^2, & -\Sigma\rho v u, & -\Sigma\rho w u, & \Sigma\rho u \\ -\Sigma\rho u v, & -\Sigma\rho v^2, & -\Sigma\rho w v, & \Sigma\rho v \\ -\Sigma\rho u w, & -\Sigma\rho v w, & -\Sigma\rho w^2, & \Sigma\rho w \\ -\Sigma\rho u, & -\Sigma\rho v, & -\Sigma\rho w, & \Sigma\rho \end{array} \quad (37{\cdot}4)$$

This tensor may be separated into two parts, the first referring to the motion, u_0, v_0, w_0, of the centre of mass of the particles in an element, and the second to their internal motions, u_1, v_1, w_1, relative to the centre of mass. With regard to the last part, $\Sigma\rho u_1 v_1$ represents the rate of transfer of u-momentum across unit area parallel to the y-plane, and is

therefore equal to the stress usually denoted by p_{xy}. Hence (37·4) becomes

$$= \begin{matrix} -p_{xx}-\rho u_0^2, & -p_{yx}-\rho v_0u_0, & -p_{zx}-\rho w_0u_0, & \rho u_0 \\ -p_{xy}-\rho u_0v_0, & -p_{yy}-\rho v_0^2, & -p_{zy}-\rho w_0v_0, & \rho v_0 \\ -p_{xz}-\rho u_0w_0, & -p_{yz}-\rho v_0w_0, & -p_{zz}-\rho w_0^2, & \rho w_0 \\ -\rho u_0, & -\rho v_0, & -\rho w_0, & \rho \end{matrix} \quad (37·5)$$

where ρ is now the whole density referred to the axes chosen.

Consider the equations

$$\frac{\partial}{\partial x_\nu}T^\nu_\mu=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (37·6)$$

Taking $\mu=4$, and using (37·5), we get the well-known equation of continuity

$$\frac{\partial(\rho u_0)}{\partial x}+\frac{\partial(\rho v_0)}{\partial y}+\frac{\partial(\rho w_0)}{\partial z}+\frac{\partial\rho}{\partial t}=0 \quad . \quad . \quad . \quad (37·7)$$

Taking $\mu=1$,

$$-\left(\frac{\partial p_{xx}}{\partial x}+\frac{\partial p_{xy}}{\partial y}+\frac{\partial p_{xz}}{\partial z}\right)=\frac{\partial(\rho u_0^2)}{\partial x}+\frac{\partial(\rho u_0v_0)}{\partial y}+\frac{\partial(\rho u_0w_0)}{\partial z}+\frac{\partial(\rho u_0)}{\partial t},$$

$$=\rho\left(u_0\frac{\partial u_0}{\partial x}+v_0\frac{\partial u_0}{\partial y}+w_0\frac{\partial u_0}{\partial z}+\frac{\partial u_0}{\partial t}\right), \text{ using (37·7)}$$

$$=\rho\frac{Du_0}{Dt}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (37·8)$$

Now (37·7) and (37·8) are the fundamental equations of hydrodynamics. By assuming Galilean axes we have neglected any extraneous body-forces, and so the term $-\rho X$, which occurs on the right side of (37·8) in the more general form of the equation, does not appear in this case.

The equation (37·6) is thus equivalent to the general equations of a fluid under no forces.

38. The equation $\partial T^\nu_\mu/\partial x_\nu=0$ represents a law of conservation. Choose one of the co-ordinates, x_4, as independent variable, and integrate the equation through a three-dimensional volume marked out in the other co-ordinates. This gives

$$\frac{\partial}{\partial x_4}\iiint T^4_\mu dx_1dx_2dx_3=-\iiint\left(\frac{\partial T^1_\mu}{\partial x_1}+\frac{\partial T^2_\mu}{\partial x_2}+\frac{\partial T^3_\mu}{\partial x_3}\right)dx_1dx_2dx_3,$$

=the surface integral of the normal component of $(T^1_\mu, T^2_\mu, T^3_\mu)$.

If the volume is such as to include the whole of the material, T^ν_μ vanishes on the surface; the surface-integral therefore

vanishes, and hence the volume integral of T^4_μ remains constant. If the surface does not include all the matter, any change of its content of T^4_μ occurs by a flux across the surface measured by $(T^1_\mu, T_\mu, T^3_\mu)$. It will be seen from (37·5) that for the axes there used T^4_μ represents the negative momentum and the mass (or energy), and that T^1_μ, &c., represent the flux of these quantities. Equation (37·6) therefore gives the law of conservation of momentum and mass, as may be verified from the corresponding hydromechanical equations.

39. Equation (37·6) is the degenerate form for Galilean co-ordinates of the covariant equation

$$T^\nu_{\mu\nu}=0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (39{\cdot}11)$$

where $T^\nu_{\mu\nu}$ is the (contracted) covariant derivative of T^ν_μ (*see* (22·7)). Equation (39·11) thus holds for Galilean co-ordinates, and it does not contain derivatives of the g's higher than the first. Hence by the principle of equivalence it holds generally, including the case of a permanent gravitational field.

Taking equation (35·8)

$$G_{\sigma\tau}-\tfrac{1}{2}g_{\sigma\tau}G=-8\pi T_{\sigma\tau},$$

multiply by $g^{\tau\nu}$. We obtain

$$G^\nu_\sigma-\tfrac{1}{2}g^\nu_\sigma G=-8\pi T^\nu_\sigma \quad . \quad . \quad . \quad . \quad (39{\cdot}12)$$

Take the covariant derivative of both sides, and contract it,

$$G^\nu_{\sigma\nu}-\tfrac{1}{2}g^\nu_\sigma\frac{\partial G}{\partial x_\nu}=-8\pi T^\nu_{\sigma\nu}=0 \quad . \quad . \quad . \quad (39{\cdot}13)$$

whence by (20·1)

$$G^\nu_{\sigma\nu}=\tfrac{1}{2}\frac{\partial G}{\partial x_\sigma} \quad . \quad . \quad . \quad . \quad . \quad . \quad (39{\cdot}14)$$

Clearly this equation will have to be an identity, and it may be verified analytically, using the values (26·3) of $G_{\mu\nu}$. For $\sigma=1, 2, 3, 4$, this identity gives the four relations between Einstein's ten equations, which have already been mentioned as reducing the number of independent conditions to six.

Conversely, from the identity (39·14) we can deduce (39·11), and hence obtain the equations of hydromechanics and the law of conservation directly from Einstein's law of gravitation. Further, by applying the hydromechanical equations to an isolated particle, we obtain the equations of motion (29). The mass of a particle has been introduced first as a constant of integration, and afterwards identified with the gravitation-mass by determining the motion of a particle in its field; it now appears that it is also the inertia-mass, because it satisfies

the law of conservation of mass and momentum, which gives the recognised definition of inertia.

It is startling to find that the whole of the dynamics of material systems is contained in the law of gravitation; at first sight gravitation seems scarcely relevant in much of our dynamics. But there is a natural explanation. A particle of matter is a singularity in the gravitational field, and its mass is the pole-strength of the singularity; consequently the laws of motion of the singularities must be contained in the field-equations, just as those of electromagnetic singularities (electrons) are contained in the electromagnetic field-equations. The fact that Einstein's law predicts these well-known properties of matter seems to be a valuable confirmation of this theory.

The general equation (39·11) enables us to pass from the equations of a fluid under no body forces to the equations of a fluid in a field of force. It can be simplified considerably. By (22·7)

$$T^{\nu}_{\mu\nu}=\frac{\partial T^{\nu}_{\mu}}{\partial x_{\nu}}-\{\nu\mu,\beta\}T^{\nu}_{\beta}+\{\beta\nu,\nu\}T^{\beta}_{\mu}. \quad . \quad . \quad (39{\cdot}21)$$

By (26·25) the last term becomes

$$\frac{1}{\sqrt{-g}}\frac{\partial}{\partial x_{\beta}}(\sqrt{-g})\,.\,T^{\beta}_{\mu}. \quad . \quad . \quad . \quad (39{\cdot}22)$$

The second term is equal to

$$-[\nu\mu,\epsilon]g^{\epsilon\beta}g^{\nu\alpha}T_{\alpha\beta}$$
$$=-\frac{1}{2}\frac{\partial g_{\nu\epsilon}}{\partial x_{\mu}}g^{\epsilon\beta}g^{\nu\alpha}T_{\alpha\beta},$$

since the other two terms cancel on summation,

$$=\frac{1}{2}\frac{\partial g^{\alpha\beta}}{\partial x_{\mu}}\,.\,T_{\alpha\beta}\,. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (39{\cdot}3)$$

This last result follows, since

$$g^{\nu\alpha}g_{\nu\epsilon}=0 \text{ or } 1,$$

so that $$g^{\nu\alpha}dg_{\nu\epsilon}+g_{\nu\epsilon}dg^{\nu\alpha}=0.$$

Multiply by $g^{\epsilon\beta}$ and use (26·15), we obtain

$$g^{\epsilon\beta}g^{\nu\alpha}dg_{\nu\epsilon}=-dg^{\alpha\beta}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (39{\cdot}4)$$

Hence inserting (39·22) and (39·3) in (39·21), we have

$$\frac{1}{\sqrt{-g}}\frac{\partial}{\partial x_{\nu}}(\sqrt{-g}\,.\,T^{\nu}_{\mu})=-\frac{1}{2}\frac{\partial g^{\alpha\beta}}{\partial x_{\mu}}\,.\,T_{\alpha\beta}\,. \quad . \quad (39{\cdot}45)$$

This equation has its simplest interpretation when we choose co-ordinates, so that $\sqrt{-g}=1$, that is to say, the volume of a four-dimensional element is to be the same in co-ordinate measure as in natural measure. Owing to the considerable freedom of choice of co-ordinates, allowed by Einstein's equations, it is always possible to do this. In that case (39·45) becomes

$$\frac{\partial}{\partial x_\nu}(T_\mu^\nu)=-\frac{1}{2}\frac{\partial g^{\alpha\beta}}{\partial x_\mu}\cdot T_{\alpha\beta}. \quad . \quad . \quad . \quad . \qquad (39·5)$$

Comparing this with (37·6), which holds when there is no field of force, we see that the term on the right represents the momentum and energy transferred from the gravitational field to the material system. As a first approximation (retaining only $T_{44}=\rho$, and $g_{44}=1-2\Omega$) we see that it gives, for $\mu=1, 2, 3$, the terms ρX, ρY, ρZ of the usual hydrodynamical equations, which were omitted in (37·8).

40. *Propagation of Gravitation.*

The velocity of light being a fundamental relation between the measures of time and space, we may expect the strains representing a varying gravitational field to be propagated with this velocity. We shall show how to derive the equations exhibiting the propagation.

In the theory of sound, the general equation of disturbances propagated with unit velocity is

$$\Box\varphi=\left(\frac{\partial^2}{\partial t^2}-\frac{\partial^2}{\partial x^2}-\frac{\partial^2}{\partial y^2}-\frac{\partial^2}{\partial z^2}\right)\varphi=\Phi, \quad . \quad . \qquad (40·11)$$

where Φ is zero except at the source of the disturbance. The general solution is

$$\varphi=\frac{1}{4\pi}\iiint\frac{\Phi'}{r'}dV' \quad . \quad . \quad . \quad . \qquad (40·12)$$

the integral being taken through the volume occupied by the source of disturbance, and the value of Φ' taken for a time $t-r'$, where r' is the distance of the volume dV' from the point considered. Thus φ is a retarded potential, and (40·12) exhibits the effect as delayed by propagation.

In the case of sound the velocity depends to a slight extent on the amplitude, and (40·11) is only strictly true if the square of φ is negligible. Similarly the velocity of light depends to a slight extent on the gravitational field (§ 32); consequently we

can only expect to obtain an equation of this form if we neglect the square of the disturbance, so that the equations become linear.

The origin of gravitational waves must be attributed to moving matter; and, since $G_{\mu\nu}$ vanishes except in a region occupied by matter, we may take $G_{\mu\nu}$ as the analogue of Φ. We shall examine whether the disturbance can be represented by a quantity $h_{\mu\nu}$ satisfying

$$\Box h_{\mu\nu} = 2G_{\mu\nu}, \quad \ldots\ldots \quad (40{\cdot}21)$$

where the exact significance of $h_{\mu\nu}$ is yet to be found. We shall regard $h_{\mu\nu}$ as a small quantity of the first order; the deviations of the $g_{\mu\nu}$ from their Galilean values will also be of the first order. Small quantities of the second order will be neglected.

If, as usual,

$$h_\mu^\sigma = g^{\nu\sigma} h_{\mu\nu},$$

and

$$h = g^{\mu\nu} h_{\mu\nu}.$$

Then, multiplying (40·21) successively by $g^{\nu\sigma}$ and $g^{\mu\nu}$, we have to this approximation,*

$$\Box h_\mu^\sigma = 2G_\mu^\sigma \quad \ldots\ldots \quad (40{\cdot}22)$$

and

$$\Box h = 2G \quad \ldots\ldots \quad (40{\cdot}23)$$

Hence

$$\Box(h_\mu^\sigma - \tfrac{1}{2}g_\mu^\sigma h) = 2(G_\mu^\sigma - \tfrac{1}{2}g_\mu^\sigma G)$$
$$= -16\pi T_\mu^\sigma \qquad \text{by (39·12).}$$

To the present approximation (37·6) holds, so that

$$\Box\left(\frac{\partial}{\partial x_\sigma} h_\mu^\sigma - \tfrac{1}{2} g_\mu^\sigma \frac{\partial h}{\partial x_\sigma}\right) = -16\pi \frac{\partial}{\partial x_\sigma} T_\mu^\sigma = 0.$$

Having regard to boundary conditions, the solution is clearly

$$\frac{\partial}{\partial x_\sigma} h_\mu^\sigma = \tfrac{1}{2} g_\mu^\sigma \frac{\partial h}{\partial x_\sigma}$$
$$= \frac{1}{2}\frac{\partial h}{\partial x_\mu} \quad \ldots\ldots\ldots \quad (40{\cdot}3)$$

* The $g^{\mu\nu}$ behave as constants until we reach equation (40·5), because their derivatives, which are small quantities of the first order, only appear in combination with the small quantities $h_{\mu\nu}$ or $G_{\mu\nu}$. The $g^{\mu\nu}$ accordingly pass freely under the differential operators.

Consider the expression

$$-\frac{1}{2}\frac{\partial}{\partial x_\alpha}\left\{g^{\alpha\beta}\left(\frac{\partial h_{\nu\beta}}{\partial x_\mu}+\frac{\partial h_{\mu\beta}}{\partial x_\nu}-\frac{\partial h_{\mu\nu}}{\partial x_\beta}\right)\right\}+\frac{1}{2}\frac{\partial}{\partial x_\mu}\left(g^{\mu\nu}\frac{\partial h_{\mu\nu}}{\partial x_\nu}\right), \quad . \quad (40\cdot4)$$

which to our approximation

$$=-\frac{1}{2}\frac{\partial^2 h^\alpha_\nu}{\partial x_\alpha \partial x_\mu}-\frac{1}{2}\frac{\partial^2 h^\alpha_\mu}{\partial x_\alpha \partial x_\nu}+\frac{1}{2}g^{\alpha\beta}\frac{\partial^2 h_{\mu\nu}}{\partial x_\alpha \partial x_\beta}+\frac{1}{2}\frac{\partial^2 h}{\partial x_\mu \partial x_\nu}.$$

By (40·3) the first two terms cancel with the last, and for Galilean values of $g^{\alpha\beta}$ the third term is simply

$$\tfrac{1}{2}\Box h_{\mu\nu}.$$

Thus by (40·21) the expression (40·4) reduces to $G_{\mu\nu}$.

Neglecting squares of small quantities, $G_{\mu\nu}$ (26·3) reduces to

$$-\frac{\partial}{\partial x_\alpha}\{\mu\nu, \alpha\}+\frac{\partial^2}{\partial x_\mu \partial x_\nu}\log\sqrt{-g}$$

$$=-\frac{1}{2}\frac{\partial}{\partial x_\alpha}\left\{g^{\alpha\beta}\left(\frac{\partial g_{\nu\beta}}{\partial x_\mu}+\frac{\partial g_{\mu\beta}}{\partial x_\nu}-\frac{\partial g_{\mu\nu}}{\partial x_\beta}\right)\right\}+\frac{1}{2}\frac{\partial}{\partial x_\mu}\left(g^{\mu\nu}\frac{\partial g_{\mu\nu}}{\partial x_\nu}\right). \quad . \quad . \quad (40\cdot5)$$

Comparing (40·4) and (40·5) we see that the h's must be equal to the g's—or rather since the h's have been treated as small quantities, they must be the deviations of the g's from their constant Galilean values. Writing $\delta_{\mu\nu}$ for the Galilean values of $g_{\mu\nu}$ (16·3), then

$$g_{\mu\nu}=\delta_{\mu\nu}+h_{\mu\nu}, \quad . \quad . \quad . \quad . \quad . \quad (40\cdot6)$$

and $h_{\mu\nu}$ satisfies the equation of wave-propagation (40·21).

By (40·12) the solution of the propagation equation is

$$h_{\mu\nu}=\frac{1}{2\pi}\iiint\frac{G'_{\mu\nu}}{r'}dV',$$

$$=-4\iiint\frac{T'_{\mu\nu}}{r'}dV'+2\delta_{\mu\nu}\iiint\frac{T'}{r'}dV'. \quad (40\cdot7)$$

This can be used for the practical calculation of $g_{\mu\nu}$ due to an arbitrary distribution of moving matter. It is necessary, as in the corresponding calculation of retarded electromagnetic potentials, to allow for the variation of $t-r'$ from point to point of the body; the boundary of dV' does not coincide with the limits of the body at any one instant. Thus for a particle of mass m, we have *

$$\iiint\frac{T'_{44}}{r'}dV'=\iiint\frac{\rho'}{r'}dV'=\frac{m}{[r(1-v_r)]},$$

* *See*, for example, Lorentz, "The Theory of Electrons," p. 254; or Cunningham, "The Principle of Relativity," p 108.

where v_r is the velocity in the direction of r, and the square bracket indicates retarded values. As is well known $[r(1-v_r)]$ is to the first order equal to the unretarded distance r, so that notwithstanding the finite velocity of propagation the force is directed approximately towards the contemporaneous position of the attracting body. It was lack of knowledge of this compensation which led Laplace and many following him to state that the velocity of gravitation must far exceed the velocity of light.

The practical application of these formulæ is, however, very limited. In a natural system (*e.g.*, the solar system) the relative velocities (u) are due to the gravitational field and u^2 is a small quantity of the first order. Consequently our approximation is not good enough to take account of T_{11}, T_{12}, &c., in natural systems; it can only include components with suffix 4.* The fact is that the whole idea of propagation from a point-source is an abstraction; actually the motion of the source, or singularity, is but the symbol of the changes occurring in all parts of the field; we cannot say whether the motion is the cause or effect of the gravitational waves.

The present solution is a particular solution. It gives unique values of the $g_{\mu\nu}$, but these may, of course, be subjected to arbitrary transformations.

* For the higher approximations needed in the problems of the solar system, see De Sitter, "Monthly Notices," Dec. 1916.

CHAPTER VII.

THE PRINCIPLE OF LEAST ACTION.

41. *Lagrange's Equations.*

We shall again restrict the choice of co-ordinates so that $\sqrt{-g}=1$. Einstein's equations (26·3) for the field in free space then becomes simplified to

$$G_{\mu\nu} \equiv -\frac{\partial}{\partial x_\alpha}\{\mu\nu, \alpha\}+\{\mu\beta, \alpha\}\{\nu\alpha, \beta\}=0. \quad . \quad (41\cdot1)$$

We shall regard $g^{\mu\nu}$ as a generalised co-ordinate (q), and x_1, x_2, x_3, x_4 as independent variables—a four-dimensional time. Writing $g^{\mu\nu}_\alpha$ for $\partial g^{\mu\nu}/\partial x_\alpha$, which will then be a generalised velocity $(\dot{q})$, we shall show that equations (41·1) can be expressed in the Lagrangian form.

$$G_{\mu\nu}=\frac{\partial}{\partial x_\alpha}\left(\frac{\partial L}{\partial g^{\mu\nu}_\alpha}\right)-\frac{\partial L}{\partial g^{\mu\nu}}=0, \quad . \quad . \quad . \quad . \quad (41\cdot2)$$

where
$$L=g^{\mu\nu}\{\mu\beta, \alpha\}\{\nu\alpha, \beta\} \quad . \quad . \quad . \quad . \quad . \quad (41\cdot3)$$

it being understood that the $g_{\mu\nu}$ are expressed as functions of the $g^{\mu\nu}$.

We have from (41·3)

$$\delta L=\{\mu\beta, \alpha\}\{\nu\alpha, \beta\}\delta g^{\mu\nu}+2g^{\mu\nu}\{\mu\beta, \alpha\}\delta\{\nu\alpha, \beta\},$$

since in the last term μ and ν are dummies.

$$=-\{\mu\beta, \alpha\}\{\nu\alpha, \beta\}\delta g^{\mu\nu}+2\{\mu\beta, \alpha\}\delta[g^{\mu\nu}\{\nu\alpha, \beta\}].$$

But

$$\delta[g^{\mu\nu}\{\nu\alpha, \beta\}]=\tfrac{1}{2}\delta\left[g^{\mu\nu}g^{\beta\lambda}\left(\frac{\partial g_{\nu\lambda}}{\partial x_\alpha}+\frac{\partial g_{\alpha\lambda}}{\partial x_\nu}-\frac{\partial g_{\alpha\nu}}{\partial x_\lambda}\right)\right].$$

The last two terms in the bracket will cancel in the summation after inner multiplication by $\{\mu\beta, \alpha\}$, because μ and β, ν and λ are interchangeable simultaneously. Also by (39·4)

$$g^{\mu\nu}g^{\beta\lambda}\frac{\partial g_{\nu\lambda}}{\partial x_\alpha}=-\frac{\partial g^{\mu\beta}}{\partial x_\alpha}.$$

Hence
$$\delta L=-\{\mu\beta, \alpha\}\{\nu\alpha, \beta\}\delta g^{\mu\nu}-\{\mu\beta, \alpha\}\delta g^{\mu\beta}_\alpha.$$

Therefore

$$\left.\begin{aligned}\frac{\partial L}{\partial g^{\mu\nu}_{\alpha}} &= -\{\mu\nu, \alpha\},\\ \frac{\partial L}{\partial g^{\mu\nu}} &= -\{\mu\beta, \alpha\}\{\nu\alpha, \beta\}\end{aligned}\right\}, \quad . \quad (41\cdot4)$$

showing that (41·1) and (41·2) are equivalent.

As in ordinary dynamics, Lagrange's equations are equivalent to

$$\int L d\tau \text{ is stationary} \quad . \quad . \quad . \quad . \quad (41\cdot5)$$

for variations of $g^{\mu\nu}$, $d\tau$ being the four-dimensional element of volume, here representing the independent variable. It must be remembered that the variations are limited by the constraint $\sqrt{-g}=1$.

42. *Principle of Least Action.**

Following out the dynamical analogy $\partial L/\partial g^{\mu\nu}_{\alpha}$ or $\partial L/\partial \dot{q}$ is to be regarded as a momentum (p). The system is dynamically of the simplest kind, since L does not contain the " time," x_{μ}, explicitly, and it is a homogeneous quadratic function of the " velocities." By the properties of homogeneous functions

$$2L = \Sigma\dot{q}\frac{\partial L}{\partial \dot{q}} = \Sigma\dot{q}p.$$

Since $(p\dot{q}+q\dot{p})$ is a perfect differential,

$$\int \Sigma(p\dot{q}+q\dot{p})d\tau$$

will be equal to a surface integral; and it will, therefore be stationary for variations of $g^{\mu\nu}$ (the variations as usual being supposed to vanish at the boundary).

Thus $$\delta\int\Sigma q\dot{p}d\tau = -\delta\int\Sigma\dot{q}pd\tau = -2\delta\int Ld\tau \quad . \quad . \quad (42\cdot1)$$

Hence, if we write

$$H = L + \Sigma q\dot{p} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (42\cdot2)$$

by (41·5) and (42·1)

$$\int H d\tau \text{ is stationary.} \quad . \quad . \quad . \quad . \quad (42\cdot3)$$

* The strict analogue of the principle of least action is the stationary property of $\int\Sigma q\dot{p}d\tau$ The restriction in dynamics that the energy is not to be varied corresponds to $\sqrt{-g}=1$. (*Cf.* §43.)

By (41·4)

$$\Sigma q\dot{p} = -g^{\mu\nu}\frac{\partial}{\partial x_\alpha}\{\mu\nu, \alpha\}.$$

Hence (42·2), (41·3) and (41·1) give

$$H = g^{\mu\nu}G_{\mu\nu} = G.$$

We can therefore write the result (42·3) thus

$$\int G \,.\, \sqrt{-g} \,.\, d\tau \quad \text{is stationary} \quad . \quad . \quad . \quad (42·4)$$

since $\sqrt{-g}=1$.

But G and $\sqrt{-g} \,.\, d\tau$ are invariants (20·3); so that (42·4) has no reference to any particular choice of co-ordinates, and the restriction $\sqrt{-g}=1$ can now be removed. It is thus a more general result than (41·5).

43. *Energy of the Gravitational Field.*

Reverting to the restriction $\sqrt{-g}=1$, multiply (41·2) by $g^{\mu\nu}_\beta$

$$g^{\mu\nu}_\beta G_{\mu\nu} = g^{\mu\nu}_\beta \frac{\partial}{\partial x_\alpha}\left(\frac{\partial L}{\partial g^{\mu\nu}_\alpha}\right) - \frac{\partial L}{\partial g^{\mu\nu}}\frac{\partial g^{\mu\nu}}{\partial x_\beta}. \quad . \quad . \quad . \quad (43·1)$$

But

$$\frac{\partial L}{\partial x_\beta} = \frac{\partial L}{\partial g^{\mu\nu}}\frac{\partial g^{\mu\nu}}{\partial x_\beta} + \frac{\partial L}{\partial g^{\mu\nu}_\alpha}\frac{\partial g^{\mu\nu}_\alpha}{\partial x_\beta}. \quad . \quad . \quad . \quad (43·2)$$

Remembering that

$$\frac{\partial}{\partial x_\beta}\, g^{\mu\nu}_\alpha = \frac{\partial^2 g^{\mu\nu}}{\partial x_\beta \partial x_\alpha} = \frac{\partial}{\partial x_\alpha}\, g^{\mu\nu}_\beta,$$

we have, adding (43·1) and (4 ·2),

$$g^{\mu\nu}_\beta G_{\mu\nu} = \frac{\partial}{\partial x_\alpha}\left(g^{\mu\nu}_\beta \frac{\partial L}{\partial g^{\mu\nu}_\alpha}\right) - \frac{\partial L}{\partial x_\beta} \quad . \quad . \quad . \quad . \quad (43·3)$$

$$= -16\pi \frac{\partial}{\partial x_\alpha}\, t^\alpha_\beta, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (43·4)$$

where

$$-16\pi\, t^\alpha_\beta = g^{\mu\nu}_\beta \frac{\partial L}{\partial g^{\mu\nu}_\alpha} - g^\alpha_\beta L. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (43·5)$$

We have used the property of g^α_β as a substitution operator.

The quantity t^α_β defined by (43·5) is the analogue of the Hamiltonian integral of energy, $\Sigma \dot{q}\frac{\partial L}{\partial \dot{q}} - L$. In free space $G_{\mu\nu}=0$, and (43·4) becomes

$$\frac{\partial}{\partial x_\alpha}\, t^\alpha_\beta = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (43·6)$$

showing that t^α_β is conserved (§ 38).

When matter is present (43·4) gives

$$-16\pi\frac{\partial t^{\alpha}_{\beta}}{\partial x_{\alpha}}=\frac{\partial g^{\mu\nu}}{\partial x_{\beta}}G_{\mu\nu},$$

$$=\frac{\partial g^{\mu\nu}}{\partial x_{\beta}}(G_{\mu\nu}-\tfrac{1}{2}g_{\mu\nu}G),$$

since, when $g=-1$, $g_{\mu\nu}dg^{\mu\nu}=0$.

Hence by (35·8)

$$\frac{\partial}{\partial x_{\alpha}}t^{\alpha}_{\beta}=\tfrac{1}{2}\frac{\partial g^{\mu\nu}}{\partial x_{\beta}}T_{\mu\nu}. \quad . \quad . \quad . \quad . \quad . \quad (43\cdot7)$$

$$=-\frac{\partial}{\partial x_{\alpha}}T^{\alpha}_{\beta} \quad \text{by (39·5).}$$

Therefore

$$\frac{\partial}{\partial x_{\alpha}}(T^{\alpha}_{\beta}+t^{\alpha}_{\beta})=0. \quad . \quad . \quad . \quad . \quad . \quad (43\cdot8)$$

This is the law of conservation in the general case when there is interaction between matter and the gravitational field. We see that the changes of energy and momentum of the matter can be regarded as due to a transfer from or to the gravitational field, the total amount being conserved. We have, in fact, traced the disappearing portion of the material tensor T^{ν}_{μ}, and shown that it reappears as the quantity t^{ν}_{μ} belonging to the gravitational field.

In order to represent the phenomena in this way we have had to restrict the choice of co-ordinates by keeping the volume of a region of space-time invariant ($\sqrt{-g}=1$). Otherwise the equation takes the more general form (39·11) which cannot immediately be interpreted as a law of conservation. It should be noted that, unlike T^{α}_{β}, the quantity t^{α}_{β} is not strictly a tensor.

44. *The Method of Hilbert and Lorentz.*

An alternative method of deriving the fundamental equations of this theory is based on the postulate that all the laws of mechanics can be summed up in a generalised principle of stationary action, viz.,

$$\delta\int(H_1+H_2+H_3+\ldots.)\sqrt{-g}\,.\,d\tau=0. \quad . \quad . \quad (44\cdot1)$$

Here H_1, H_2, H_3 are invariants* involving, respectively, the parameters describing the gravitational field, the electromagnetic field, and the material system. If we consider

* Invariant because the equation must hold in all systems of co-ordinates, and we already know that the factor $\sqrt{-g}\,.\,d\tau$ is invariant.

matter and radiation in bulk we may add a fourth term involving the entropy, so as to bring in thermodynamical phenomena, and so on. The variations are taken with respect to these parameters, their values at the boundary of integration being kept constant.

It is well known that the laws of mechanics of matter and of electrodynamics can be expressed in this form, so that we are here chiefly concerned with H_1. We already know from (42·4) that Einstein's theory is given by $H_1=G$. Now G is, in fact, the principle invariant of the quadratic form $g_{\mu\nu}dx_\mu dx_\nu$, viz., the Gaussian invariant of curvature. This aspect of the theory seems to eliminate any element of arbitrariness which may have been felt when we fixed on the contracted Riemann-Christoffel tensor for the law of gravitation.

To interpret G as a curvature, consider a surface drawn in space of five dimensions, whose equation referred to the lines of curvature and the normal (z) at a point on it may be written

$$2z=k_1x_1^2+k_2x_2^2+k_3x_3^2+k_4x_4^2+\text{higher powers} \quad . \quad (44·2)$$

where k_1, k_2, k_3, k_4 are the reciprocals of the principle radii of curvature.

Then $$ds^2=dz^2+\Sigma dx_1^2.$$

Eliminating z by (44·2)

$$ds^2=(1+k_1^2x_1^2)dx_1^2+\ \ldots\ldots\ +2k_1k_2x_1x_2dx_1dx_2+\ldots \quad (44·3)$$

Hence at the origin,

$$g_{\mu\mu}=1, \qquad g_{\mu\nu}=0\ (\mu\neq\nu), \qquad \partial g_{\mu\nu}/\partial x_\sigma=0.$$

The only surviving terms in $G=g^{\mu\nu}G_{\mu\nu}$ are

$$-g^{\mu\mu}\frac{\partial}{\partial x_\rho}\{\mu\mu,\ \rho\}+g^{\mu\mu}\frac{\partial^2}{\partial x_\mu^2}(\log\sqrt{-g}).$$

We easily find that

$$G=-2(k_1k_2+k_2k_3+k_3k_1+k_1k_4+k_2k_4+k_3k_4) \quad . \quad (44·4)$$

In three dimensions we have only two curvatures, and k_1k_2 is known as Gauss's measure of curvature, *i.e.*, the ratio of the solid angle contained by the normals round the perimeter of an element to the area of the element. The expression (44·4) is a generalisation of this invariant to five dimensions.

The curvature G in ordinary matter is quite considerable. In water the curvature is the same as that of a spherical space of radius 570,000,000 km. Presumably, if a globe of water of this radius existed, there would not be room in space for anything else.

45. *Electromagnetic Equations.*

The electromagnetic field is described by a covariant vector $\varkappa_\mu$. In Galilean co-ordinates,

$$\varkappa_\mu = (-F, -G, -H, \Phi), \quad . \quad . \quad . \quad (45\cdot1)$$

where F, G, H is the vector potential, and Φ the scalar potential of the ordinary theory.

If $\varkappa_{\mu\nu}$ is the covariant derivative of $\varkappa_\mu$, we have by (22·2)

$$\begin{aligned} \frac{\partial \varkappa_\mu}{\partial x_\nu} - \frac{\partial \varkappa_\nu}{\partial x_\mu} &= \varkappa_{\mu\nu} \\ &= F_{\mu\nu}, \text{ say} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (45\cdot2) \end{aligned}$$

The electric and magnetic forces are given in the electromagnetic theory by

$$X = -\frac{\partial \Phi}{\partial x} - \frac{\partial F}{\partial t}, \qquad a = \frac{\partial H}{\partial y} - \frac{\partial G}{\partial z}, \quad . \quad . \quad . \quad (45\cdot3)$$

i.e.,

$$X = \frac{\partial \varkappa_1}{\partial x_4} - \frac{\partial \varkappa_4}{\partial x_1}, \qquad a = \frac{\partial \varkappa_2}{\partial x_3} - \frac{\partial \varkappa_3}{\partial x_2}.$$

Hence by (45·2) the value of $F_{\mu\nu}$ in Galilean co-ordinates is

$$\begin{array}{c} \mu \\ \downarrow \\ \nu \end{array} \quad F_{\mu\nu} = \begin{array}{cccc} 0 & -\gamma & \beta & -X \\ \gamma & 0 & -a & -Y \\ -\beta & a & 0 & -Z \\ X & Y & Z & 0 \end{array} \quad . \quad . \quad . \quad (45\cdot41)$$

and the associated contravariant tensor, $F^{\mu\nu} = g^{\mu\alpha} g^{\nu\beta} F_{\alpha\beta}$, is

$$F^{\mu\nu} = \begin{array}{cccc} 0 & -\gamma & \beta & X \\ \gamma & 0 & -a & Y \\ -\beta & a & 0 & Z \\ -X & -Y & -Z & 0 \end{array} \quad . \quad . \quad . \quad (45\cdot42)$$

We can now express Maxwell's equations in covariant form. In the ordinary theory they are

$$\frac{\partial Z}{\partial y} - \frac{\partial Y}{\partial z} = -\frac{\partial a}{\partial t}, \quad \frac{\partial X}{\partial z} - \frac{\partial Z}{\partial x} = -\frac{\partial \beta}{\partial t}, \quad \frac{\partial Y}{\partial x} - \frac{\partial X}{\partial y} = -\frac{\partial \gamma}{\partial t}, \quad (45\cdot51)$$

$$\frac{\partial \gamma}{\partial y} - \frac{\partial \beta}{\partial z} = \frac{\partial X}{\partial t} + u, \quad \frac{\partial a}{\partial z} - \frac{\partial \gamma}{\partial x} = \frac{\partial Y}{\partial t} + v, \quad \frac{\partial \beta}{\partial x} - \frac{\partial a}{\partial y} = \frac{\partial Z}{\partial t} + w, \quad (45\cdot52)$$

$$\frac{\partial X}{\partial x} + \frac{\partial Y}{\partial y} + \frac{\partial Z}{\partial z} = \rho, \quad . \quad . \quad . \quad . \quad . \quad . \quad (45\cdot53)$$

$$\frac{\partial a}{\partial x} + \frac{\partial \beta}{\partial y} + \frac{\partial \gamma}{\partial z} = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (45\cdot54)$$

where the velocity of light is unity, and the Heaviside-Lorentz unit of charge is chosen so that the factor 4π disappears. The

electric current u, v, w and the density of electric charge ρ form a contravariant vector, since

$$(u, v, w, \rho)=\Sigma e\left(\frac{dx}{ds}, \frac{dy}{ds}, \frac{dz}{ds}, \frac{dt}{ds}\right) \text{ per unit volume,*}$$

$$=J^{\mu}, \text{ say} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (45{\cdot}6)$$

Equations (45·51) and (45·54) may be written,

$$\frac{\partial F_{\mu\nu}}{\partial x_\sigma}+\frac{\partial F_{\nu\sigma}}{\partial x_\mu}+\frac{\partial F_{\sigma\mu}}{\partial x_\nu}=0, \quad . \quad . \quad . \quad . \quad (45{\cdot}71)$$

and the remaining equations (45·52) and (45·53) give

$$\frac{\partial F^{\mu\nu}}{\partial x_\nu}=J^{\mu}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (45{\cdot}72)$$

Now (45·71) is satisfied identically on substituting the values of $F_{\mu\nu}$ from (45·2), so that (45·2) and (45·72) represent the fundamental electromagnetic equations. The former is already covariant, and the latter is made covariant by writing the covariant derivative for the ordinary derivative. Thus

$$F^{\mu\nu}_{\nu}=J^{\mu}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (45{\cdot}81)$$

$$F_{\mu\nu}=\frac{\partial \varkappa_\mu}{\partial x_\nu}-\frac{\partial \varkappa_\nu}{\partial x_\mu}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (45{\cdot}82)$$

are the required equations. These hold in the gravitational field because the conditions for the application of the principle of equivalence (§ 27) are satisfied.

The expression $F^{\mu\nu}_{\nu}$ may be simplified as in § 39 ; but owing to the antisymmetry of $F^{\mu\nu}$ the term corresponding to (39·3) disappears, and the equation reduces to

$$\frac{1}{\sqrt{-g}}\frac{\partial}{\partial x_\nu}(\sqrt{-g}\,.\,F^{\mu\nu})=J^{\mu}. \quad . \quad . \quad . \quad (45{\cdot}9)$$

The fact that Maxwell's equations can be reduced to a covariant form shows that all electromagnetic phenomena described by them will be in agreement with the principle of relativity.

* The occurrence of ds instead of dt in the denominator is due to the Michelson-Morley contraction, $\beta=dt/ds$, which makes the estimate of unit volume by a fixed observer differ from that made by an observer moving with the electrons. (*Cf.* equation (7·65).)

46. *The Electromagnetic Energy-Tensor.*

According to the electromagnetic theory, the components of mechanical force on unit volume containing electric charges are

$$k_1 = \rho X + \gamma v - \beta w,$$
$$k_2 = \rho Y + \alpha w - \gamma u,$$
$$k_3 = \rho Z + \beta u - \alpha v,$$

and the negative rate of doing work is

$$k_4 = -Xu - Yv - Zw,$$

since the magnetic force does no work.

By (45·41) and (45·6), these give

$$\begin{aligned} -k_\nu &= F_{\mu\nu} J^\mu \\ &= F_{\mu\nu} F^{\mu\sigma}_{\sigma}, \end{aligned} \qquad \ldots \quad (46{\cdot}1)$$

so that k_ν is a vector.

But k_ν represents the rate at which the momentum and negative energy of the material system are being increased, *i.e.*, in Galilean co-ordinates,

$$-\frac{\partial}{\partial x_\alpha} T^\alpha_\nu = k_\nu. \qquad \ldots \quad (46{\cdot}2)$$

If there exists a corresponding tensor E^α_ν for the electromagnetic field, this must change by an equivalent amount in the opposite direction in order to satisfy the law of conservation. Thus

$$\frac{\partial}{\partial x_\alpha} E^\alpha_\nu = k_\nu. \qquad \ldots \quad (46{\cdot}3)$$

It is not difficult to show from (46·1) and (46·3) that

$$E^\alpha_\nu = -F_{\nu\beta} F^{\alpha\beta} + \tfrac{1}{4} g^\alpha_\nu F^{\sigma\tau} F_{\sigma\tau}. \qquad \ldots \quad (46{\cdot}4)$$

We omit the proof as the precise value is not of great interest to us. It is sufficient to know that the expression is of the necessary tensor-form, so that an energy-tensor for the electromagnetic field exists.

In general co-ordinates (46·2) and (46·3) are replaced by the covariant equations,

$$-T^\alpha_{\nu\alpha} = k_\nu = E^\alpha_{\nu\alpha} \qquad \ldots \quad (46{\cdot}5)$$

in accordance with the principle of equivalence.

When no matter is present this gives $E^\alpha_{\nu\alpha} = 0$, and we can derive the reaction of the gravitational field just as in (39·5). It follows that electromagnetic energy in the gravitational field experiences a force just as material energy does. Further

electromagnetic energy exerts gravitation, because (39·13) and (46·5) give

$$(G^a_\nu - \tfrac{1}{2}g^a_\nu G)_a = 0 = -8\pi(T^a_\nu + E^a_\nu)_a,$$

the lower a denoting covariant differentiation.

Hence on integrating, (39·12) must be replaced by

$$G^a_\nu - \tfrac{1}{2}g^a_\nu G = -8\pi(T^a_\nu + E^a_\nu).$$

In fact the electromagnetic energy-tensor must simply be added on to the material energy-tensor throughout our work.

When $\sqrt{-g}=1$, we have the most general law of conservation for triangular interchanges between matter, electromagnetism and gravitation.

$$\frac{\partial}{\partial x_a}(T^a_\nu + E^a_\nu + t^a_\nu) = 0. \quad . \quad . \quad . \quad . \quad (46·6)$$

47. *The Aether.*

The application of the Calculus of Variations to (44·1) gives a number of differential equations equal to the number of parameters varied ; but, according to a general theorem due to Hilbert, there are always four identical relations between these equations (the number 4 corresponding to the dimensions of $d\tau$). The number of independent equations is thus four less than the number of unknowns, so that in addition to arbitrary boundary conditions we can impose four arbitrary relations on the parameters. It is this freedom of choice of co-ordinates that is so fundamental a characteristic of the generalised principle of relativity.

If we vary H_1 only we find the ten equations $G_{\mu\nu}=0$. The identical relations in this case have been given in § 39. If we vary the electromagnetic variable $\varkappa_\mu$ as well, we get 14 equations, of which 10 are independent, to determine 14 unknowns Within certain limits we can give arbitrary values to four of the unknowns, and the other ten will then be determined definitely by the equations and the boundary conditions. If we elect to fix the values of the four co-ordinates $\varkappa_\mu$ in this way (so that they are, as it were, disposed of) the $g_{\mu\nu}$ will become fixed, that is to say, there will be only one possible space-time. The phenomena, electromagnetic as well as gravitational, will all be described by the $g_{\mu\nu}$, which represent the state of strain of this space-time. This space-time may be materialised as the aether, and the aether-theory does in fact attribute electromagnetic phenomena to strains in this supposed absolute medium.

This is only a crude indication of the relation of the aether-theory to our relativity theory. As is well known, the modern aether-theory involves rotational strains. Moreover, we cannot get rid of the electromagnetic variables by putting them equal to zero, because they form a vector, which cannot vanish in one system of co-ordinates without vanishing in all.

48. *Summary of the Last Two Chapters.*—It may be useful to review the results which have been obtained from the point at which we introduced the energy-tensor T_μ^ν of the material system. Initially it was brought in for the practical purpose of calculating the gravitational field of a material body; but this has led on to a discussion of the general laws of dynamics.

As mentioned in § 6, it is important, if we wish to adopt the principle of relativity, to show that the laws of nature which we generally accept are consistent with the principle; or, if not, to modify them so that they may become consistent. We have had to modify one law—the law of gravitation. The laws of mechanics (Newton's laws of motion) are equivalent to the conservation of momentum and the conservation of mass. We have in § 7(*c*) found it necessary to generalise the latter by admitting that energy has mass; and the conservation of mass is absorbed in the conservation of energy. The most general statement of these two principles of conservation for material systems is found in the general equations of hydrodynamics (or of the theory of gases), viz., (37·7) and (37·8), and it is therefore sufficient to verify these. We have done that by showing that they may be expressed in tensor-form. We have even gone further; we have shown that these laws can actually be deduced from the law of gravitation. They correspond to the four identical relations between Einstein's ten equations of gravitation (§ 39).

It has similarly been verified that our electromagnetic equations are of tensor-form and are therefore consistent with relativity. But in this case we have not deduced the electromagnetic equations from anything else; we have merely shown their admissibility. The energy-tensor E_μ^ν of the electromagnetic field is found from the consideration that in interchanges between the material and electromagnetic systems the total momentum and energy must remain constant.

When the co-ordinates are not Galilean, gravitational forces will be acting and the total energy and momentum of the material and electromagnetic systems will be altering. We

have shown how to find this flux of energy and momentum (39·5), and in § 43 we have traced it into the gravitational field, showing that it reappears there as the quantity t_μ^ν, which, moreover, is conserved when no transfer of this kind is going on. There is, however, one reservation necessary; unlike T_μ^ν and E_μ^ν, t_μ^ν is not a tensor, and in order that this complete conservation of energy and momentum may be apparent we have to choose co-ordinates so that $\sqrt{-g}=1$. This does not imply any exception to the physical law of conservation, because we can always choose co-ordinates satisfying this condition. *It is merely that the energy-tensor is slightly* more general than the physical idea of energy and momentum; the former may be reckoned with respect to any co-ordinates, the latter must be reckoned with respect to co-ordinates satisfying $\sqrt{-g}=1$.

From the existence of an energy-tensor for the electromagnetic field, it is deduced that electromagnetic energy must experience and exert gravitational force.

The remainder of our work has been principally concerned with showing that our equations are equivalent to a principle of least action. From a theoretical standpoint there is a great deal to be said in favour of reversing the whole procedure, starting from the principle of least action as a postulate; but I have preferred the present course as more elementary.

Some difficulty may be found in the fact that the time-component of a four-dimensional vector is usually called by a different name from the space-components. The following table may be useful for reference :—

Vector.	*Space-Components.*	*Time-Component.*
T_μ^4	negative momentum	energy (mass).
T_μ^1	flux of negative momentum	flux of energy (mass).
k_μ	force	negative rate of doing work.
$\varkappa_\mu$	negative vector potential...	electric scalar potential.
J^μ	electric current-density ...	electric charge-density.

CHAPTER VIII.

THE CURVATURE OF SPACE AND TIME.

49. We have now presented the laws of gravitation, of hydromechanics, and of electromagnetism, in a form which regards all systems of co-ordinates as on an equal footing. And yet it is scarcely true to say that all systems are equally fundamental; at least we can discriminate between them in a way which the restricted principle of relativity would not tolerate.

Imagine the earth to be covered with impervious cloud. By the gyro-compass we can find two spots on it called the Poles, and by Foucault's pendulum-experiment we can determine an angular velocity about the axis through the Poles, which is usually called the earth's absolute rotation. The name "absolute rotation" may be criticised; but, at any rate, it is a name given to something which can be accurately measured. On the other hand, we fail completely in any attempt to determine a corresponding "absolute translation" of the earth. It is not a question of applying the right name—there is no measured quantity to name. It is clear that the equivalence of systems of axes in relative rotation is in some way less complete than the equivalence of axes having different translations; and this may perhaps be regarded as a failure to reach the ideals of a philosophical principle of relativity.

This limitation has its practical aspect. We might suppose that from the expression (28·8) for the field of a particle at rest it would be possible by a transformation of co-ordinates to deduce the field of a particle, say, in uniform circular motion. But this is not the case. We may, of course, reduce the particle to rest by using rotating axes; but we find it necessary to take an entirely different solution of the partial differential equations, satisfying different boundary conditions.

We have not hitherto paid any attention to the invariance of the boundary conditions; and it is here that the breakdown occurs. The axes ordinarily used in dynamics are such

that as we recede towards infinity in space the $g_{\mu\nu}$ approach the special set of values (16·3). On transforming to other co-ordinates the differential equations are unaltered; but usually the boundary values of the $g_{\mu\nu}$, and consequently the appropriate solutions of the equations, are altered. We can, therefore, discriminate between different systems of co-ordinates according to the boundary values of the g's; and *those which* at infinity pass into Galilean co-ordinates may properly be considered the most fundamental, since the boundary values are most simple. The complete relativity for uniform translation is due to the boundary values as well as the differential equations remaining unaltered.*

We have based our theory on two axioms—the restricted principle of relativity and the principle of equivalence. These taken together may be called the *physical* principle of relativity. We have justified, or explained, them by reference to a *philosophical* principle of relativity, which asserts that experience is concerned only with the relations of objects to one another and to the observer and not to the fictitious space-time framework in which we instinctively locate them. We are now led into a dilemma; we can save this philosophical principle only by undermining its practical application. The measurement of the rotation of the earth detects something of the nature of a fundamental frame of reference—at least in the part of space accessible to observation. We shall call this the "inertial frame." Its existence does not necessarily contradict the philosophical principle, because it may, for instance, be determined by the general distribution of matter in the universe; that is to say, we may be detecting by our experiments relations to matter not generally recognised. But having recognised the existence of the inertial frame, the philosophical principle of relativity becomes arbitrary in its application. It cannot foretell that the Michelson-Morley experiment will fail to detect uniform motion relative to this frame; nor does it explain why the acceleration of the earth relative to this frame is irrelevant, but the rotation of the earth is important.

The inertial-frame may be attributed (1) to unobserved world-matter, (2) to the aether, (3) to some absolute character

* Owing to the four additional conditions that can be imposed on the g's the boundary values are not sufficient to determine the co-ordinates uniquely and the principle of relativity is valid in its most complete sense for transformations considerably more general than uniform translation.

of space-time. It is doubtful whether the discrimination between these alternatives is more than word-splitting, but they lead to rather different points of view. The last alternative seems to contradict the philosophical principle of relativity, but in the light of what has been said the physicist has no particular interest in preserving the philosophical principle. In this chapter we shall consider two suggestions towards a theory of the inertial frame made by Einstein and de Sitter respectively. These should be regarded as independent speculations, arising out of, but not required by, the theory hitherto described.

The inertial frame is distinguished by the property that the $g_{\mu\nu}$ referred to it approach the limiting Galilean values (16·3) as we recede to a great distance from all attracting matter. This is verified experimentally with considerable accuracy; but it does not follow that we can extrapolate to distances as yet unplumbed, or to infinity. If it is assumed that the Galilean values still hold at infinite distances, the inertial frame is virtually ascribed to conditions at infinity, and its explanation is removed beyond the scope of physical theory. We may, however, suppose that observational results relate to only a minute part of the whole world, and that at vaster distances the $g_{\mu\nu}$ tend to zero values which would be invariant for all finite transformations. In that case all frames of reference are alike at infinity, and the property of the inertial frame arises from conditions within a finite distance. In that case physical theories of the inertial frame may be developed.

The ascription of the inertial frame to boundary conditions at infinity may also be avoided by abolishing the boundary. This is really only another aspect of the vanishing of the $g_{\mu\nu}$ at infinity. Our four-dimensional space-time may be regarded as a closed surface in a five-dimensional continuum; it will then be unbounded but finite, just as the surface of a sphere is unbounded.

We have seen (§44) that wherever matter exists space-time has a curvature. It might seem that if there were sufficient matter the continuum would curve round until it closed up; but it has not been found possible to eliminate the boundary so simply. I think the difficulty arises because time is not symmetrical with respect to the other co-ordinates; in general matter moves with small velocity, so that the different components of the energy-tensor T_μ^ν are not of the same order of magnitude.

50. Einstein suggests that in measurements on a vast scale the line-element has the form

$$ds^2 = -R^2\{d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta d\phi^2)\} + dt^2. \quad . \quad (50\cdot1)$$

This expression includes the effects of the general distribution of matter through space; but there will be superposed the local irregularities due to its condensation into stellar systems, etc.

The expression (50·1) can be interpreted* as belonging to a three-dimensional space which forms the surface of a hypersphere of radius R in four dimensions, the time being rectilinear. Let O be the origin of co-ordinates (Fig. 5), A the

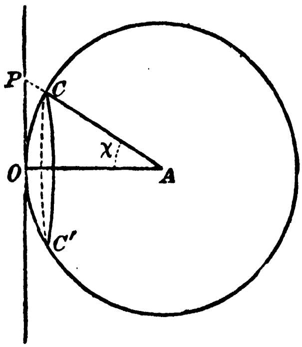

FIG. 5.

centre of the hypersphere, and χ the angle OAC. If θ is the azimuthal angle of the plane OAC, the line-element at C for an ordinary sphere would be

$$-ds^2 = R^2 d\chi^2 + R^2 \sin^2\chi d\theta^2.$$

The expression (50·1) is the extension of this for an extra dimension measured by ϕ.

In the figure the circumference of the circle CC' is $2\pi R\sin\chi$, but its radius measured along the sphere is $R\chi$. Similarly in our curved space the surface of a sphere of radius $R\chi$ will be $4\pi R^2\sin^2\chi$; successively more distant spheres will increase in

* Other interpretations are possible; but this is probably the easiest conception for those unfamiliar with non-Euclidean geometry. For this reason I do not here describe the interpretation in terms of "elliptical space," which has certain advantages.

area up to a radius $\frac{1}{2}\pi R$, and afterwards decrease to a point for the limiting distance πR. The whole volume of space is finite and equal to $2\pi^2 R^3$ in natural measure.*

From (50·1) the values of $G_{\mu\nu}$ can be calculated just as in § 28. We find, in fact,

$$\left.\begin{aligned} G_{\mu\nu} &= \frac{2}{R^2} g_{\mu\nu}, \quad \text{except } G_{44}=0 \\ \text{so that} \qquad G &= \frac{6}{R^2} \end{aligned}\right\} \quad . \quad . \quad . \quad (50{\cdot}2)$$

Hence by (35·8)

$$-8\pi T_{\mu\nu} = -\frac{1}{R^2} g_{\mu\nu}, \text{ except } -8\pi T_{44} = -\frac{3}{R^2}. \quad . \quad (50{\cdot}3)$$

Unless we are willing to suppose that the matter in the universe is moving with speeds approaching that of light, T_{44} is much greater than the other components, and it is clearly impossible to satisfy (50·3). The only possible course is to make a slight modification of the law of gravitation. Neglecting the motion of matter we shall have $T_{44}=\rho$, and the other components vanish. The modified law that satisfies (50·2) must then be

$$G_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu}(G-2\lambda) = -8\pi T_{\mu\nu} \quad . \quad . \quad . \quad (50{\cdot}4)$$

where $\lambda = 1/R^2$ and $\rho = 1/4\pi R^2$.

Equation (50·4) replaces (35·8). The radius R may be as great as we please, so that we may satisfy our scruples without introducing any modification perceptible to observation.

In Hamilton's principle G becomes replaced by $G-4\lambda$, and space-time has a natural curvature 4λ when no matter is present; this curvature is increased to 6λ where there is matter having the average density. (*Cf.* (44·4) with $k_4=0$.)

Since the whole volume of space in natural measure is $2\pi^2 R^3$, the total mass of matter is $2\pi^2 R^3 \rho = \frac{1}{2}\pi R$. The mass of the sun is 1·47 kilometres; the mass of the stellar system may be estimated at $10^9 \times$ sun; let us suppose further that the spiral nebulæ represent 1,000,000 stellar systems having this mass. Even this total mass will only give us a universe of radius 10^{15} kilometres, or about 30 parsecs—much less than the average distance of the naked-eye stars. Einstein's hypothesis therefore demands the existence of vast quantities of undetected matter which we may call world-matter.

* The observer will probably introduce measures more convenient to himself (*cf.* § 52), so that in his co-ordinates the limiting distance may be ∞ or even beyond.

Some curious results are obtained by following out the properties of this spherical space. The parallax of a star diminishes to zero as the distance (in natural measure) increases up to $\frac{1}{2}\pi R$; it then becomes negative and reaches $-90°$ at a distance πR. Apart from absorption of light in space we should see an anti-sun, at the point of the sky opposite to the sun equally large and equally bright,* the surface-markings corresponding to the back of the sun. After travelling "round the world" the sun's rays come back to a focus. Since ρ and R are related, it has been suggested that we can use the invisibility of this anti-sun to give a lower limit to R, assuming that no light is lost in space except by the scattering action of the world-matter. But it appears to have been overlooked that Einstein's new hypothesis is inconsistent with relativity in its ordinary sense; the anti-sun will not be a virtual image of the sun as it is now, but of the sun as it was when it emitted the light—perhaps millions of years ago, when it was in another part of the stellar system. Einstein has restored the differentiation between space and time by assuming the space-time world to be cylindrical, so that the linear direction gives an absolute time. It is only locally that we can still make Minkowski's transformation; rigorously the physical principle of relativity is violated since space-time is no longer isotropic.

We regret being unable to recommend this rather picturesque theory of anti-suns and anti-stars. It suggests that only a certain proportion of the visible stars are material bodies; the remainder are ghosts of stars, haunting the places where stars used to be in a far-off past.

Owing to this violation of the restricted principle of relativity we have a feeling that Einstein's new hypothesis throws away the substance for the shadow. It is also open to the serious criticism that the law of gravitation is made to involve a constant λ, which depends on the total amount of matter in the universe ($\lambda=\pi^2/4M^2$). This seems scarcely conceivable; and it looks as though the solution involves a very artificial adjustment.

51. An alternative proposal has been made by de Sitter which seems much less open to objection. He takes for the line element

$$ds^2=-R^2\{d\chi^2+\sin^2\chi(d\theta^2+\sin^2\theta d\varphi^2)\}+\cos^2\chi dt^2. \quad (51\cdot1)$$

For constant time the three-dimensional space is spherical as in (50·1); but there is also a curvature in the time-variable.

* Disregarding the sun's absolute motion referred to later.

With the present variables this is not of a simple kind, but setting

$$\left.\begin{aligned} \sin\chi &= \sin\zeta\sin\omega \\ \tan(it/R) &= \cos\zeta\tan\omega \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (51\cdot2)$$

we find

$$ds^2 = -R^2(d\omega^2 + \sin^2\omega(d\zeta^2 + \sin^2\zeta(d\theta^2 + \sin^2\theta d\varphi^2))). \quad (51\cdot3)$$

which corresponds to spherical polar co-ordinates $(R, \omega, \zeta, \theta, \varphi)$ in space of five dimensions. By measuring ζ from different azimuths we perform an operation corresponding to Minkowski's rotation of the time-axis, so that there is here no absolute time, and the original principle of relativity is fully satisfied.

The properties of de Sitter's space-time are best recognised from (51·1). Near the origin we have ordinary Galilean space-time. As we recede, space has the spherical properties already mentioned, and in addition measured time (ds) begins to run slow relative to co-ordinate time (dt). Finally at $\chi = \frac{1}{2}\pi$, *i.e.*, at a natural distance $\frac{1}{2}\pi R$, time stands still. At any fixed point ds is zero however large dt may be, so that nothing whatever can happen however long we wait.

Of course, this is merely the point of view of the observer at the origin of co-ordinates. All parts of this spherical continuum are interchangeable; and if our observer could transport himself to this peaceful abode, he would find Nature there as active as ever. Moreover, adopting the co-ordinates natural to his new position, he would judge his old home to be in this passive state. There is a complete lack of correspondence between the times at the two places. They are, as it were, at right angles, so that the progress of time at one point has no relation to the perception of time at the other point.

The line-element (51·1) leads to

$$G_{\mu\nu} = \frac{3}{R^2} g_{\mu\nu},$$

and accordingly the law of gravitation is taken to be (50·4), with

$$\lambda = 3/R^2.$$

The aggregate curvature due to matter is here neglected in comparison with the natural curvature due to the modification of the law of gravitation, and there is no assumption of the existence of vast quantities of matter not yet recognised.

There is no anti-sun on de Sitter's hypothesis, because light, like everything else, is reduced to rest at the zone where time stands still, and it can never get round the world. The region beyond the distance $\frac{1}{2}\pi R$ is altogether shut off from us by this barrier of time. The parallax of a star at this distance will be such as corresponds to a distance R in Euclidean space, and this is the minimum value possible.

The most interesting application of this hypothesis is in connection with the very large observed velocities of spiral nebulæ, which are believed to be distant sidereal systems. Since $\sqrt{g_{44}}=\cos\chi$, the vibrations of the atoms become slower (in the observer's time) as $\cos\chi$ diminishes, in accordance with § 34. We should thus expect the spectral lines to be displaced towards the red in very distant objects, an effect which would in practice be attributed to a great velocity of recession. It is not possible to say as yet whether the spiral nebulæ show a systematic recession, but so far as determined up to the present receding nebulæ seem to preponderate.

Superposed on the (spurious) systematic radial velocity will be the individual velocities of the nebulæ. It is scarcely possible to say what these are likely to be without making some assumption. There is no meaning in absolute motion, and if two systems are entirely independent, so that their relative motion has no physical cause, it must be quite arbitrary, and there is no reason to expect it to be small compared with the velocity of light. If, however, the systems have separated from one another, it can be shown by rather laborious calculations* that their velocities will tend to become more diverse as they recede, up to the limit $\frac{1}{4}\pi R$ for which the velocities are comparable with that of light. We should thus have an explanation of the large velocities of the spirals, averaging 300-400 km. per sec., and we could perhaps form an estimate of the value of R.

It must be remembered that in natural measure the internal motions of stars in a spiral system will be of the same magnitude as in our own system, owing to the homogeneous character of de Sitter's space-time. In co-ordinate measure these internal motions will be smaller owing to the transformation of the time. The possibility of large divergent motions of the systems as a whole depends on the large separation between them.

* De Sitter, "Monthly Notices," November, 1917.

52. So far we have used spherical co-ordinates, but we can map the spherical space of Einstein or of de Sitter on a flat space by performing the central projection $r=R\tan\chi$ r will be represented by OP in Fig. 5, and the variables r, θ, ϕ. will satisfy Euclidean geometry. This does not mean that measured space is Euclidean; but that we multiply our measures by suitable factors in order to obtain results which will fit together in Euclidean space, just as we did for a local gravitational field in § 28. With r as variable (50·1) and (51·1) become, respectively,

$$ds^2=\frac{-dr^2}{(1+\varepsilon r^2)^2}-\frac{r^2}{1+\varepsilon r^2}(d\theta^2+\sin^2\theta d\phi^2)+dt^2 \quad . \quad (52\cdot1)$$

$$ds^2=\frac{-dr^2}{(1+\varepsilon r^2)^2}-\frac{r^2}{1+\varepsilon r^2}(d\theta^2+\sin^2\theta d\phi^2)+\frac{dt^2}{1+\varepsilon r^2} \quad (52\cdot2)$$

where $\varepsilon=1/R^2$.

These show that at "infinity" (*i.e.*, $r=\infty$) the values of $g_{\mu\nu}$ in rectangular co-ordinates approach the respective limits.

EINSTEIN.				DE SITTER.				GALILEO.			
0	0	0	0	0	0	0	0	−1	0	0	0
0	0	0	0	0	0	0	0	0	−1	0	0
0	0	0	0	0	0	0	0	0	0	−1	0
0	0	0	1	0	0	0	0	0	0	0	1

the Galilean values being added for comparison.

De Sitter's limiting values are invariant for all transformations; Einstein's only for transformations not involving the time; the Galilean values for the transformation of uniform motion and a limited group of other transformations.

De Sitter's hypothesis thus appears to present the greatest advantages; but it will not satisfy the followers of Mach's philosophy. He derives his inertial-frame from the spherical property of space-time which in turn is derived from the slightly modified law of gravitation; it is not determined by anything material. The followers of Mach maintain that if there were no matter there could be no inertial frame, and it appears that this is Einstein's reason for preferring his own suggestion. In his theory if all matter were abolished, R would become zero and the world would vanish to a point. There is something rather fascinating in a theory of space by which, the more matter there is, the more room is provided. It is satisfactory, too, from Einstein's standpoint, because he is unwilling to admit that a thinkable space without matter

could exist. For our part, we feel equally unwilling to assent to the introduction of vast quantities of world-matter, which (to quote de Sitter) "fulfils no other purpose than to enable us to suppose it not to exist."

53. In this discussion of the law of gravitation, we have not sought, and we have not reached, any ultimate explanation of its cause. A certain connection between the gravitational field and the measurement of space has been postulated, but this throws light rather on the nature of our measurements than on gravitation itself. The relativity theory is indifferent to hypotheses as to the nature of gravitation, just as it is indifferent to hypotheses as to matter and light. We do not in these days seek to explain the behaviour of natural forces in terms of a mechanical model having the familiar characteristics of matter in bulk; we have to accept some mathematical expression as an axiomatic property which cannot be further analysed. But I do not think we have reached this stage in the case of gravitation.

There are three fundamental constants of nature which stand out pre-eminently—

The velocity of light, $3{\cdot}00 . 10^{10}$ C.G.S. units ; dimensions LT^{-1}.
The quantum, $6{\cdot}55 . 10^{-27}$ „ ; „ ML^2T^{-1}.
The constant of gravitation, $6{\cdot}66 . 10^{-8}$ „ ; „ $M^{-1}L^3T^{-2}$.

From these we can construct a fundamental unit of length whose value is

$$4 \times 10^{-33} \text{ cms.}$$

There are other natural units of length—the radii of the positive and negative unit electric charges—but these are of an altogether higher order of magnitude.

With the possible exception of Osborne Reynolds's theory of matter, no theory has attempted to reach such fine-grainedness. But it is evident that this length must be the key to some essential structure. It may not be an unattainable hope that some day a clearer knowledge of the processes of gravitation may be reached; and the extreme generality and detachment of the relativity theory may be illuminated by the particular study of a precise mechanism.